BIOLOGIA PLANTARUM 41 (3): 339-355, 1998

Quick identification of the members of the glutamine synthetase gene family from sunflower by simultaneous amplification of cDNA with degenerate primers

M. MONTENEGRO, J. M. MALDONADO and R. PÉREZ-VICENTE*

Departamento de Biología Vegetal, División de Fisiología Vegetal, Facultad de Ciencias, Universidad de Córdoba, E-14071 Córdoba, Spain

Abstract

A single degenerate glutamine synthetase (GS)-specific primer was used to amplify the 3' end of cDNAs derived from different GS genes that are expressed in leaves and roots of sunflower (*Helianthus annuus* L. cv. Peredovic). Four types of GS cDNA (I, II, III and IV) were simultaneously amplified from leaves and five types (I, II, V, VI, VII) from roots with a minimum investment of time and experimental work. cDNAs II, III and IV encode chloroplastic isoforms as deduced by the presence of chloroplastic GS-specific features in their sequences. The rest of cDNAs codifies cytosolic isoforms. Using cDNA-specific probes and primers, homologous sequences to all GS cDNAs amplified from cv. Peredovic, except to cDNAs III and IV, were detected in the inbred line R41. This result strongly suggests that the three cDNAs for chloroplastic isoform are allelic sequences from the same locus, and since cDNA type IV contains sequences derived from cDNAs II and III, it indicates a recombinational origin. The results presented are consistent with the existence of a GS gene family in sunflower with at least five members. Four of them, named *ggs*1.1 to *ggs*1.4, codify for the cytosolic isoforms (cDNAs I, V, VI and VII). A fifth member, named *ggs*2, from which three allelic sequences (cDNAs II, III and IV) have been cloned, encodes the chloroplastic isoform.

Additional key words: chloroplastic isoforms, cytosolic isoforms, *Helianthus annuus*.

Received 17 April 1998, *accepted* 20 August 1998.
Acknowledgements: This research was supported by Dirección General de Investigación Científica y Técnica (grant no. PB93-0718), Dirección General de Enseñanza Superior (PB96-0504-C02-01) and Junta de Andalucía (PAI, group no. 3261), Spain. The authors express their sincere thanks to Dr. G. Dorado and Dr. M. Ruiz for their critical readings.
* Corresponding author, fax: (+95) 7 218606, e-mail: bv1pevir@lucano.uco.es

Introduction

Glutamine synthetase (GS, EC 6.3.2.1) is the enzyme that catalyses the ATP-dependent synthesis of glutamine from glutamate and ammonium. In plants, GS exists as multiple isoforms that are responsible for the assimilation of the ammonium released by distant processes (Miflin and Lea 1980). The subunits of the GS isoforms are encoded by small families of nuclear genes whose expression is differentially regulated (Forde and Cullimore 1989, Lam *et al.* 1996). The gene for the chloroplastic isoform is induced by light and exhibits enhanced expression in photosynthetic cells, in good agreement with the main task of its product, the reassimilation of photorespired ammonium (Wallsgrove *et al.* 1987, Edwards and Coruzzi 1989, Lam *et al.* 1996). Different number of genes for the cytosolic isoforms have been reported depending on the species considered, showing distinct patterns of expression that support the preferential participation of their product in particular processes (Forde and Cullimore 1989, Lam *et al.* 1996). Genes for cytosolic isoforms that are preferentially expressed in cotyledons during germination and in senescing cotyledons have been reported in pea and radish (Coruzzi *et al.* 1991, Watanabe *et al.* 1994). Various legumes contain cytosolic GS genes with enhanced expression in N_2-fixing nodules (Bennett *et al.* 1989, Walker and Coruzzi 1989, Roche *et al.* 1993, Temple *et al.* 1995). In maize, the treatment with ammonium or nitrate increases the amount of transcripts of GS1-1 in roots and of GS1-2 in shoots, whilst a third gene for the cytosolic isoforms of GS (GS1-3) is unaffected by those nitrogen compounds (Sukanya and Snustad 1994).

Consistently with this background, any approach to the understanding of the role of GS in the metabolism of a new species will require to determine the number and identity of the operating GS genes as well as the regulation of their expression. In practical terms, this imply the cloning and characterisation of multiple sequences. To date, plant GS genes have been cloned by traditional time-consuming methods based in the preparation and screening of libraries with heterologous probes. This partially explains why, while GS-coding sequences have been cloned from numerous plants, the characterization of complete GS gene families has been achieved in relatively few cases. As an alternative to traditional methods of cloning, GS-specific degenerate primers have been designed that drive the amplification of GS-coding sequences from a wide spectrum of plants (Pérez-Vicente *et al.* 1996).

Sunflower is an important crop plant that raises interesting questions for basic and applied research, many of which have not been approached in molecular terms. The identity and location of the GS isoforms in sunflower have been determined (Cabello *et al.* 1991) and regulation studies have been performed at the level of enzyme activity. Light, nitrate and the stage of development differentially regulate the activity of the GS isoforms of sunflower cotyledons along germination (de la Haba *et al.* 1992). A control at the level of gene transcription is expected to contribute to the observed changes in enzyme activity (Forde and Cullimore 1989, Lam *et al.* 1996). The importance of this regulatory step is ignored since, to date, the number and identity of the sunflower GS genes remain unknown.

This work was aimed to determine the number and type of GS genes operating in sunflower. A single gene for the chloroplastic isoform and four genes for cytosolic isoforms have been identified by analysing partial cDNAs. Unique regions of sequence for the preparation of gene-specific probes have been obtained. The cDNAs were cloned in a context of reduced time and experimental simplicity by means of PCR amplification with degenerate primers.

Materials and methods

Plant cultivation and DNA/RNA isolation: Sunflower plants (*Helianthus annuus* L.) were grown in plastic trays containing a 1:1 (v/v) mixture of perlite and vermiculite, under a 16-h photoperiod with irradiance of 200 μmol(photon) m^{-2} s^{-1} (400-700 nm) provided by *Sylvania cool white F72T12/CW/VHO*, 160 W fluorescent lamps supplemented with *Mazda* 60W incandescent bulbs) and a light/dark temperature of 25/19 °C and relative humidity of 70/80 %. Germinated seedlings were irrigated daily with a nutrient solution with 10 mM KNO_3 (Hewitt 1966). Seeds from the sunflower cv. Peredovic were donated by *Eurosemillas* (Córdoba, Spain). The inbred line R41 was a generous gift from Dr. J. Domínguez, Centro de Investigación y Formación Agraria, Córdoba, Spain. Young leaves and roots were collected from 10 - 15 individual 20-d-old plants, and total RNA and genomic DNA were isolated from them as described previously (Pérez-Vicente *et al.* 1996).

cDNA synthesis and amplification: Single-stranded cDNA, with an anchor sequence at the 5' end, was prepared by priming 5 μg of total RNA from leaves and roots with the Q_T oligonucleotide (ccagtgagcagagtgacgaggactcgagctcaagcttttttttttttttttt) described by Frohman (1994). Reactions were performed by using the M-MLV retrotranscriptase (*Promega*, Madison, USA) according to Yang *et al.* (1989). The PCR amplifications were carried out with the *GeneAmp PCR Core* kit (*Perkin-Elmer*, Norwalk, USA) in a final volume of 50 mm^3. Each reaction mix contained: 1 unit of *Taq* DNA polymerase, 200 μM of each of the four dNTPs, 2 mM of $MgCl_2$, 1 μM of each gs5 (ggagaggttatgcctggicartg) and QoR (ccagtgagcagagtgacgagga) primers (Pérez-Vicente *et al.* 1996) and 1 mm^3 of the cDNA solution, in 10 mM Tris-HCl / 50 mM KCl, pH 8.3, buffer. Primers were isolated from the rest of the reagents in the vials by means of a wax barrier to avoid false priming during the first steps of the reaction, according to the 'hot start' technique (Frohman *et al.* 1988, Chou *et al.* 1992). The PCR reactions were performed in a DNA *Thermal Cycler 480* (*Perkin-Elmer*, Norwalk, USA) using the following 40 cycle profile: 94 °C, 60 s + 60 °C, 30 s + 72 °C, 60 s plus a 6 s extension per cycle.

Specific amplification of cDNAs II, III and IV was performed under the same conditions described before, except that QoR was replaced by the corresponding cDNA-specific primer and that the annealing temperature was increased to 70 °C. Specificity of the primers was confirmed by using plasmids carrying cDNA II, III and IV inserts as target.

Cloning and restriction analysis of the amplified gs5-QoR fragments: The products of amplification with gs5 and QoR were blunt ended by treatment with the Klenow enzyme (Sambrook *et al.* 1989) and separated by electrophoresis. The major bands showed in Fig. 1*A* (lanes L and R) were excised from the gel, purified with the *Qiaex* gel extraction kit (*Qiagen*, Chatsworth, USA) and cloned into the *Eco*R V site of p*Bluescript* SK (+) (*Stratagene*, La Jolla, USA) by standard methods (Sambrook *et al.* 1989). Sixty randomly collected clones from roots and leaves were selected for restriction analysis. The enzymes *Eco*R V, *Hae* III, *Alu* I, *Hinc* II, *Sac* I and *Xba* I from *Boehringer* (Mannheim, Germany) were used to digest both the crude PCR products and the selected gs5-QoR clones. Crude PCR products (5 mm^3) were treated for 16 h at 37 °C with 15 units of each enzyme to ensure a complete digestion. The gs5-QoR inserts from individual clones were PCR amplified and digested according to the instruction of the manufacturer. The result of the digestion were electrophoresed through 2 % (m/v) *Metaphor* agarose (*FMC*, Rockland, USA) and visualized by staining with ethidium bromide.

Sequencing and sequence analysis: The cDNA inserts in plasmid vectors were sequenced by the chain termination method, using the *AmpliTaq DyeDeoxy Terminator Cycle Sequencing* kit (*Perkin-Elmer/Applied Biosystems*, Foster City, USA). Two to three independent clones of each cDNA type (I to VII) were sequenced on both directions to avoid PCR-induced mutations. The nucleotide sequence data obtained will appear in the *GenBank Nucleotide Sequence Database* under the accession numbers AF004847 (cDNA I), AF005031 (cDNA V), AF005032 (cDNA VI), AF005222 (cDNA VII), AF005223 (cDNA II), AF005224 (cDNA III) and AF005225 (cDNA IV).

Computer applications used for sequence analysis were in the *Wisconsin Package Version 9.0* from *Genetics Computer Group* (Madison, USA). *PILEUP* was used to align sequences and to obtain the corresponding dendrogram. Sequence identity values from pair-wise comparisons were calculated with *OLDISTANCES*. Sequence alignments that show the common positions shadowed were displayed with *PRETTYBOX* and those that show a consensus sequence were obtained with *PRETTY*.

Design of cDNA-specific primers: cDNA-specific primers were designed with the help of the *Oligo* application (version 4.0.5, 1994) from *National Biosciences* (Plymouth, USA). Primers specific for cDNAs II (gagctttcaaagagtgtacgcatca), III (tgtggcgaagtgaaaggttcaaa) and IV (gacgcagtttggaccaagcac) were designed from unique regions located in the 3' untranslated region and the coding region as indicated in Fig. 6. Primers specific for the cDNAs I (cccctccatattgttattcaaaataga), V (agccaagttcatgcccaactagt), VI (aaaccaaccatttacgttgttgttat) and VII (tgttgtcactgatagtctagcgacgta) were designed from cDNA-specific sequences located at the end of the 3'untranslated regions.

Synthesis of cDNA-specific probes and hybridization experiments: Probes specific for cDNAs I, V, VI and VII were obtained from their 3' untranslated regions. Fragments

containing the 3' untranslated regions were excised from cDNA I and cDNA V with *Rsa* I (restriction site at 70 and 10 nt after the stop codon, respectively), from cDNA VI with *Aci* I (restriction site at 26 nt after the stop codon) and from cDNA VII with *Sty* I (restriction site at 11 nt after the stop codon). Single-stranded probes were then prepared from the isolated 3' untranslated regions by PCR with [α^{32}P]-dATP and the corresponding cDNA-specific primer as described by Konat *et al.* (1994).

Twenty microgrammes of genomic DNA from isogenic line R41 were slot blotted onto *Hybond-N* filters (*Amersham*, Little Chalfont, UK) following the instructions of the manufacturer. Filters were hybridized with radioactively labeled cDNA-specific probes. Hybridizations were performed at 68 °C and filters were washed three times, 15 min each, in 0.2 × SSC at 68 °C following standard procedures (Sambrook *et al.* 1989). Specificity of the probes and hybridization conditions were previously tested by hybridization to plasmid carrying the cDNA I, V, VI and VII inserts.

Results and discussion

Simultaneous amplification of cDNAs coding for different GS isoforms: In the latest years, the polymerase chain reaction (PCR) has become an advantageous technique to clone cDNA mainly because its superior sensitivity compared to traditional methods. However, the use of PCR to clone plant GS sequences has been limited to the isolation of one or, at most, two different cDNAs from each species selected (Pérez-Rodriguez and Valpuesta 1996, Pérez-Vicente *et al.* 1996), while the potentiality of this technique to characterize a complete GS gene family remains untested.

We have used a cloning strategy that exploits the capacity of degenerate primers to amplify diverse sequences that share conserved regions, in order to simplify the experimental manipulation required to clone the multiple members of a gene family, while keeping a high sensitivity of sequence screening.

A fragment corresponding to the 3' half of the mRNA for the GS was selected for amplification. This fragment spanned *ca.* 750 bp and contained part of the coding region as well as the 3' untranslated sequence. Amplification was driven by primers QoR and gs5. QoR is complementary to the sequence of the oligonucleotide Q_T that is anchored to the cDNAs used as amplification target (Fig. 1*B*). gs5 is a GS-specific degenerate primer designed to match a wide variety of plant GS cDNAs without restrictions regarding the type of isoform encoded (Pérez-Vicente *et al.* 1996). Under these conditions, the simultaneous amplification of various cDNAs derived from the different GS genes that are expressed in the selected organ is expected (Fig. 1*A*). In each case, a band of the expected size was the mayor PCR product. Since the cDNAs for plant GS exhibit a remarkable size homogeneity (Forde and Cullimore 1989, Pérez-Vicente *et al.* 1996), the amplification of a single band does not imply the amplification of a single cDNA. Several cDNAs for GS of similar size but different sequence can be grouped together in a wide band (Fig. 1*A*). To check this possibility, the amplification product was treated with various restriction enzymes. Some of the restriction patterns obtained indicated the existence of more than a single type of

cDNA in the amplification product. For instance, when the product of amplification of leaves cDNA was treated with *Eco*R V, only part of the cDNAs were digested (Fig. 2). In other cases, the addition of the sizes of the fragments derived from the digestion exceeded the size of the digested amplification product, thus indicating sequence diversity (digestion with *Alu* I and *Hae* III in Fig. 2).

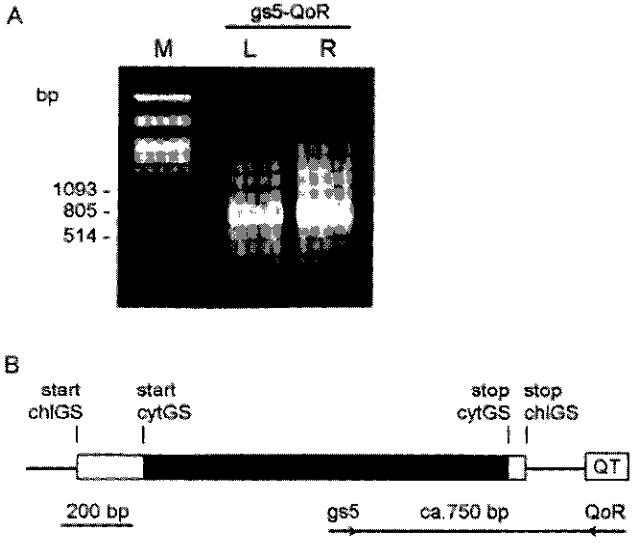

Fig. 1. PCR amplification of cDNAs coding for GS. *A.* Samples of cDNA from sunflower leaves (L) and roots (R) were amplified with QoR, a primer for the anchored oligo Q_T, and gs5, a GS-specific degenerate primer. The products of the amplification were electrophoresed in a 0.8 % (m/v) agarose gel and stained with ethidium bromide. 0.005 cm^3 from each PCR reaction (total 50 mm^3) were loaded per lane. M - molecular size marker corresponding to 0.5 μg of λ phage DNA digested with *Pst* I. *B.* Model of the cDNA for cytosolic and chloroplastic isoforms of GS with indication of the position and expected length of the fragment amplified with gs5 and QoR. The coding region of the cytosolic subunit (*black box*) is showed overlapping the longer coding region of the chloroplastic subunit (empty box). The 5' and 3' untranslated regions are representes by lines.

To isolate the different cDNAs amplified simultaneously, the amplification products were cloned in plasmids and identified by their restriction patterns. As expected, two types of clones that differ in the presence of a restriction site for *Eco*R V were identified in the product of amplification of leaves cDNA (clones 31 and 46 in Fig. 2). In the same sample, three types of cDNAs were discriminated by using *Alu* I (clones 2, 4 and 5 in Fig. 2), that, at the same time, belonged to one of the two types defined by *Eco*R V. The cDNA amplified from sunflower leaves and roots was further analysed with *Hae* III, *Hinc* II, *Sac* I and *Xba* I. A total of nine types of cDNA (I to IX) were identified. cDNAs III and IV and cDNAs V to IX were exclusively found in leaves and roots, respectively. cDNAs I and II were ubiquitous. Clones carrying the same type of cDNA of different length were also identified by this method (clones 10, 12 and 13 in Fig. 2). Length differences were due to

polyadenylation at alternative sites (Fig. 3), a common feature among plant mRNAs (Dean *et al.* 1986). The existence of the AAUAAA motive preceding the polyadenylation site was not a tight requirement for the polyadenylation of sunflower mRNAs (Fig. 3), as observed in many other plant species (Li and Hunt 1997).

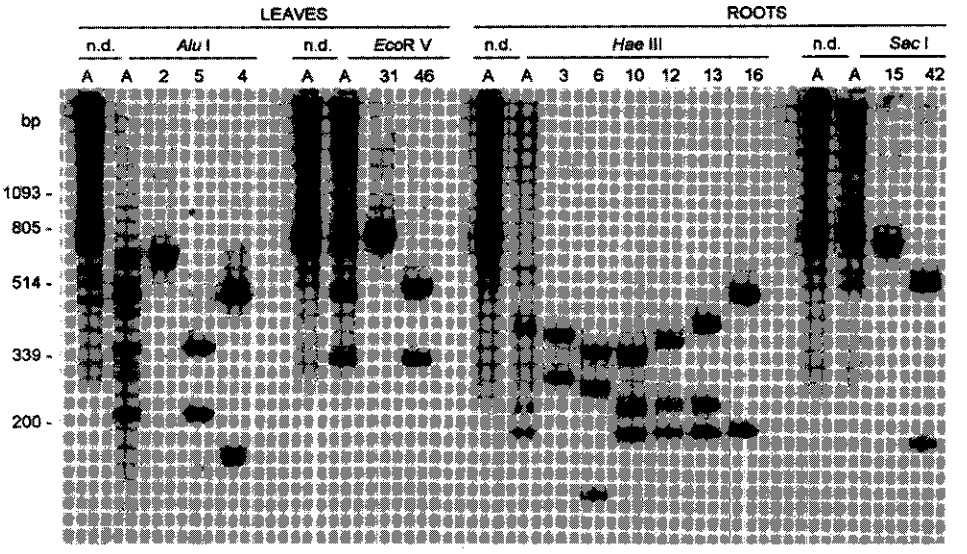

Fig. 2. Identification of different cDNAs simultaneously amplified with gs5 and QoR. The crude PCR products (A) and the individual gs5-QoR fragments isolated from them were treated with the indicated enzymes as described in Materials and Methods or remained undigested (n.d.). The restrictions fragments were separated by electrophoresis in 2 % (m/v) agarose gels and stained with ethidium bromide. Gel image is shown in negative. DNA size markers are indicated on the left.

```
501  AtgaAATTAT GAAGCCTGAT TTGGAGGGAG ATTGAAAAAA TTTAATTGGA

551  ATTTGAAAAC AACCCTTCTA TATGTACTTT GTTAAAATTT AGGTGCATCC

601  TACTACTCTG TTTCCTTGTT TATTTTTACC TGATGTTTTA TATGTTCTTG

651  ATGCAAAGCA TTTTGGGTCC AAAATGCTAA TGGCAACTAC TTGATTGCAT

                10                  12
701  TTGAGTTT  TCCCATTGGGAATATTGTC  AATATTTCTATTTTGAATAACAA

                          13
751  TATGGAGGGGTTGATTTAA  GATTAAAAAAA AAAAAAAAAA
```

Fig. 3. Multiple sites of polyadenylation on the mRNA for sunflower GS. Figure shows the 3' end of the GS cDNA type I, as obtained from clone h17. Alternative sites of polyadenylation on this cDNA are indicated by the number of the particular clones where they were found (10, 12 and 13). A putative polyadenylation signal is underlined. The stop codon is typed in lower case.

The sequences of the coding region of all cDNA identified, except VIII and IX, were closely related to those of the genes that encode cytosolic and chloroplastic GS isoforms from *Phaseolus* and other higher plants. The average amino acid sequence identity among the cDNAs from sunflower and the GS sequences from *Phaseolus* was 84 %, and the lowest value obtained for a particular comparison was 74 % (Table 1). As expected from sequences that are not under a severe selection pressure, the degree of conservation of the 3' untranslated regions was low. The average nucleotide sequence identity of this region within the cDNAs from sunflower was comparable to that found among the cDNAs from sunflower and *Phaseolus* (31 %). As an exception, cDNAs II, III and IV exhibited a high sequence identity in the 3' untranslated regions ranging from 66 to 87 % (Table 1).

The cDNA VIII contained sequences that matched both primers gs5 and QoR, even though it did not codify for any GS. The cDNA IX displayed QoR sequences at both ends. The presence of false positives was expected since, given the high yield and apparent specificity of the products of the first PCR (Fig. 1*A*), no reamplification with nested primers was performed. They appeared in low frequency (1 out of 60 clones analysed in both cases), thus not being a serious impediment for the methodology employed.

Identification of the cDNAs coding for cytosolic and chloroplastic isoforms of GS: The most obvious feature of the mRNAs for chloroplastic isoforms, the region that encodes the transit peptide that facilitates the entrance to the chloroplast, is placed at the 5'end of the messenger (Forde and Cullimore 1989). However, other distinctive features can be found at the 3' end or are distributed along the sequence.

The 3' end of the coding region from cDNAs II, III and IV was extended about 15 amino acids further than the rest of the sunflower cDNAs (Fig. 4). This 3' extension, that was also present in the cDNA for the chloroplastic GS from *Phaseolus*, is an exclusive feature of the messengers that code for chloroplastic GS (Forde and Cullimore 1989, Pérez-Vicente *et al.* 1996), thus indicating that cDNAs II, III and IV code for chloroplastic isoforms while the rest of cDNAs, that lack such 3' extension, encode cytosolic isoforms.

In fact, it was found that a stretch as short as the first 250 nucleotides of each cDNA (not including the 3' extension), contained enough information to determine the type of coded isoform. The dendrogram obtained from the comparison of those fragments clearly differentiated two groups of cDNAs. One group included the cDNA for the chloroplastic GS of *Phaseolus* and the sunflower cDNAs that displayed the 3' extension. The second group was formed by the cDNA for the cytosolic GS of *Phaseolus* and the rest of the cDNAs from sunflower (Fig. 5).

Four short regions of sequence conserved in the GS of mammals, plants and prokaryotes are involved in the formation of the active site of the enzyme (Eisenberg *et al.* 1987, Forde and Cullimore 1989). GS cDNAs amplified from sunflower comprised regions 2, 3 and 4 (Fig. 4). Region 1 was 5' upstream from the amplified fragments. Region 2 coincided with the binding site for primer gs5 (not shown in Fig. 4).

Table 1. Pair-wise comparison of the cDNAs for GS amplified from sunflower. Numerals above the diagonal correspond to the amino acid sequence identity of the coding region of each pair of cDNAs showed in Fig. 4. Numerals below the diagonal show the nucleotide sequence identity of the 3' untranslated regions. The highest values found are boxed. The corresponding sequences of the cDNAs for the cytosolic (beta) and chloroplastic (delta) GS from *Phaseolus* have been included as a reference. The values of sequence identity were calculated as the number of conserved positions divided by the length of the shorter sequence.

	II	IV	III	**delta**	VI	VII	I	**beta**	V	
II		0.97	0.96	**0.91**	0.83	0.81	0.77	**0.81**	0.72	II
IV	0.68		0.94	**0.89**	0.82	0.80	0.76	**0.79**	0.71	IV
III	0.66	0.87		**0.89**	0.80	0.78	0.74	**0.77**	0.70	III
delta	**0.35**	**0.33**	**0.34**		**0.87**	**0.84**	**0.80**	**0.81**	**0.74**	**delta**
VI	0.33	0.26	0.26	**0.27**		0.96	0.93	**0.87**	0.86	VI
VII	0.35	0.28	0.37	**0.33**	0.33		0.93	**0.87**	0.86	VII
I	0.42	0.27	0.39	**0.36**	0.40	0.36		**0.89**	0.85	I
beta	**0.35**	**0.22**	**0.31**	**0.29**	**0.33**	**0.34**	**0.30**		**0.81**	**beta**
V	0.38	0.34	0.31	**0.24**	0.42	0.32	0.48	**0.34**		V

```
          1                                                  50
II        GFQVGPSVGI EAGDHIWCAR YLLERITEQA GVVLTLDPKP IEGDWNGAGc
IV        GFQVGPSVGI DSGDHIWRAR YLLERITEQA GVVLTLDPKP IEGDWNGAGc
III       GSQVGPSVGI EAGDHIWCAR YLLERITEQA GVVLTLDPKP IEGDWNGAGc
delta     EYQVGPSVGI EAGDHIWASR YILERITEQA GVVLSLDPKP IEGDWNGAGc
VI        EFQVGPSVGI LAGDEIWAAR YILERITEIA GVVVSFDPKP IPGDWNGAGA
VII       EFQVGPSVGI SAGDEIWAAR YILERITEIA GVVVSFDPKP IPGDWNGAGA
I         EFQVGPSVGI SAGDELWVAR YILERITEIA GVVVSFDPKP IPGDWNGAGA
beta      GFQVGPAVGI SAGDELWVAR YILERITEVA GVVLSFDPKP IKGDWNGAGA
V         EFQIGPSVGI SAADEFWVAR YILERVTEIA GVVVSFDPKP vPGDWNGAGg
Con       --Q-GP-VGI ---D--W--R Y-LER-TE-A GVV---DPKP --GDWNGAG-
                                                        3
          51                                                 100
II        HTNYSTKAMR EDGGFEVIKK AILNLSLRHT EHISAYGEGN ERRLTGKHET
IV        HTNYSTKAMR EDGGFEVIKK AILNLSLRHT EHISAYGEGN ERRLTGKHET
III       HTNYSTKAMR EDGGFEVIKK AILNLSLRHT EHISAYGEGN ERRLTGKHGT
delta     HTNYSTKSMR EDGGFEVIKK AILNLSLRHK EHISAYGEGN ERRLTGKHET
VI        HTNYSTKSMR EEGGYEIIKK AIEKLGLRHK EHIAAYGEGN ERRLTGKHET
VII       HTNYSTKSMR EEGGYEIIKK AIEKLGLRHS EHIAAYGEGN ERRLTGRHET
I         HTNYSTKSMR DEGGYEVIKK AIEKMGLKHK EHIAAYGEGN ERRLTGRHET
beta      HTNYSTKTMR NDGGYEEIKS AIQKLGKRHK EHIAAYGEGN ERRLTGRHET
V         HTNYSTKSMR EDGGYETIQK AIEKLGLRHE DHIAGYGEGN ERRLTGHHET
Con       HTNYSTK-MR --GG-E-I-- AI------H- -HI--YGEGN ERRLTG-H-T

          101                                                150
II        ASINQFSWGV ANRGcSIRVG RDTEKAGKGY LEDRRPASNM DPYTVTGLLA
IV        ASINQFSWGV ANRGcSIRVG GDTEKAGKGY LEDRRPASNM DPYTVTGLLA
III       ASINQFSWGV ANRGcSIRVG RDTDAAGKGY LEDRRPASNM DPYTVTGLLA
delta     ASINTFSWGV ANRGcSIRVG RDTEKNGKGY LEDRRPASNM DPYVVTSLLA
VI        ANINTFKWGV ANRGASIRVG RDTEKDGKGY FEDRRPASNM DPYVVTSMIA
VII       ANINTSKWGV ANRGASIRVG RDTEKEGKGY FEDRRPASNM DPYVVTSMIA
I         ADINTSLWGV ANRGASIRVG RDTEKEGKGY FEDRRPASNM DPYVVTSMIA
beta      ADINTFLWGV ANRGASIRVG RDTEKAGKGY FEDRRPASNM DPYVVTSMIA
V         ADINTSSWGV AkRGvSIRVG RDTAEEGKGY FEDRRPGSNM DPYVVTSMIA
Con       A-IN---WGV A-RG-SIRVG -DT---GKGY -EDRRP-SNM DPY-VT---A
                     4
          151                  175
II        ETTILWEPTL EAEALAAQKL ALNV*
IV        ETTILWEPSL EAEALAAQKL ALNV*
III       ETPFLWEPSL EAEALAAQKL ALNV*
delta     ESTLLWEPTL EAEALAAQKL ALKV*
VI        ETTILL*... .......... .....
VII       ETTILYVNP* .......... .....
I         ETTIFVEQIL KIIRFGP*.. .....
beta      DTTILWKP*. .......... .....
V         ETTIIVEAFR LFCGSE*... .....
Con       ---------- ---------- -----
```

Fig. 4. Comparison of the sequences from the coding region of the sunflower GS cDNAs. Amino acid sequences of cDNAs type I to VII (from the end of primer gs5 to the stop codons) are shown aligned to the corresponding regions of the chloroplastic (delta) and cytosolic (beta) GS from *Phaseolus*. Regions involved in the formation of the active site of the enzyme are underlined and variant amino acids in them are in lower case. The 3' extensions exclusive of the chloroplastic isoforms are boxed. Consensus sequence (Con) shows positions of unanimous agreement only.

The presence of a cysteine residue in regions 3 and 4 is an exclusive feature of the chloroplastic isoforms (Forde and Cullimore 1989). Such cysteine residues occurred

exclusively in cDNAs that showed an extended coding region (II, III and IV from sunflower and delta from *Phaseolus*), thus confirming these cDNAs as chloroplastic isoform-coding sequences.

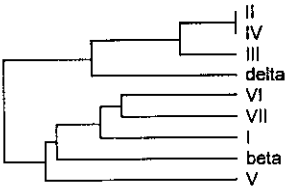

Fig. 5. Dendrogram of relationships among the GS cDNAs from sunflower and *Phaseolus*. The dendrogram was derived from the comparison of the first 250 nucleotides of each GS cDNA from sunflower and the corresponding sequences of the cDNAs for the cytosolic (beta) and chloroplastic (delta) GS from *Phaseolus*. Distances along the horizontal axis are proportional to the differences between sequences.

Occurrence of a recombination between the sequences that codify for chloroplastic isoforms: cDNAs II, III and IV, identified as for chloroplastic isoforms, are a singular group of sequences exhibiting the highest sequence identity in both the coding region and the 3' untranslated sequences of all cDNAs isolated from sunflower (Table 1, Fig. 4). Their presence in sunflower contradicts the general observation of a single gene for the chloroplastic isoform of the GS in each plant species (Forde and Cullimore 1989, Lam *et al.* 1996).

The comparison of the sequences of cDNAs II, III and IV showed that one of them (IV) was composed from sequences derived from the other two. The first half of cDNA IV (nucleotides 1 to 460 in Fig. 6) displayed very few differences when compared to that of cDNA II, while its second half (nucleotides 476 to 700) was almost identical to the corresponding region of cDNA III. Between the first and the second half there was an intermediate region of 15 nucleotides where all three cDNA were identical (Fig. 6). The most probable explanation for this situation is that cDNA IV had been originated by a recombination between cDNAs II and III, one of the crossing-over points being placed in the intermediate region. The second crossing-over point might have happened at different distances from the first one and his location is unknown by now, neither the corresponding reciprocal product of the recombination has been cloned. After recombination, the resulting new gene seems to have diverged form the parental ones as deduced from the accumulation of point differences in its sequence (Fig. 6).

The occurrence of a recombination strongly suggests that cDNAs II, III and IV represent allelic sequences from a single locus. The most frequent type of recombination in higher plants takes place during the pairing of homologous chromosomes in the meiosis, a mechanism that involves the alignment of allelic sequences (Watson *et al.* 1987, Berg and Singer 1992). Since cv. Peredovic studied was not a certified inbred line and the material for RNA extraction was collected from different individuals, the possibility of having cloned allelic sequences can not be discarded.

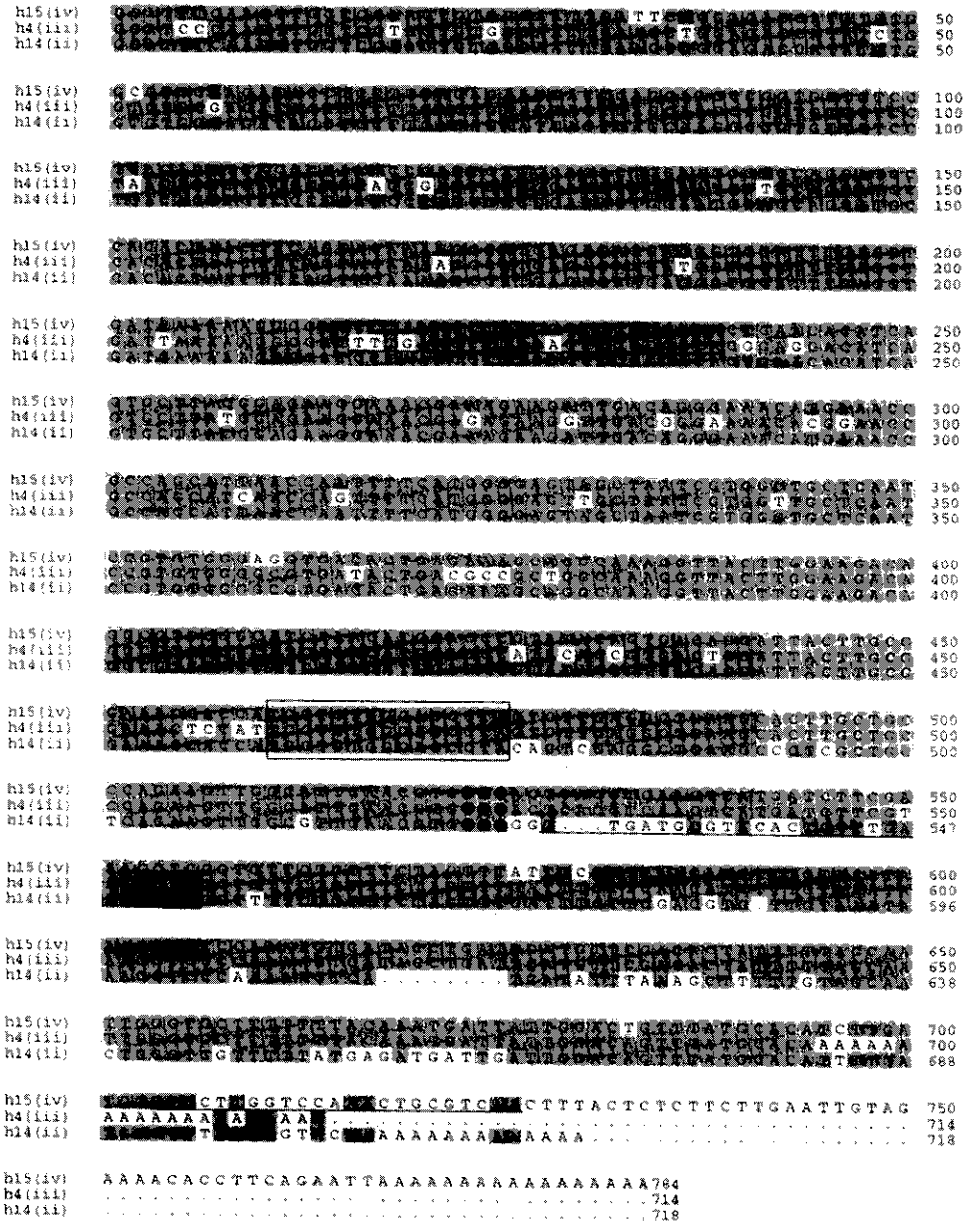

Fig. 6. Alignment of the nucleotide sequences of the GS cDNAs from sunflower encoding chloroplastic isoforms. Sequences aligned encompass part of the coding region and the 3' untranslated sequences. Conserved positions are shadowed. Nucleotides from the stop codons are hidden by black dots. The region of recombination between cDNA II and III to form cDNA IV is boxed. cDNA-specific primers were designed from the underlined sequences.

Although recombination is a well known mechanism for the generation of genetic variability, its participation in the creation of new variant of GS genes has not been reported yet. As a result of intra-allelic recombination as the one observed in sunflower, previously existing genes may interchange stretches of their sequences to generate novel GS subunits. Therefore, different polypeptidic domains that have evolved separately may be assembled in novel advantageous combinations.

The sunflower GS gene family: Taken together, the results obtained are consistent with the existence of a multigene family with at least five members. Four genes named *ggs*1.1, *ggs*1.2, *ggs*1.3 and *ggs*1.4 (cDNAs I, V, VI and VII, respectively) encode cytosolic isoforms of the GS. The mRNAs were isolated mostly from roots and lack the 3' extension. A fifth gene named *ggs*2 encodes the chloroplastic isoform, from which three putative allelic sequences (a, b and c) have been cloned (cDNAs II, III and IV, respectively) (Table 2).

Table 2. The GS gene family from sunflower.

Locus	Allele	Encoded isoform	cDNA	cDNA origin
*ggs*1.1		cytosolic	I	leaves, roots
*ggs*2	a	chloroplastic	II	leaves, roots
	b	chloroplastic	III	leaves
	c	chloroplastic	IV	leaves
*ggs*1.2		cytosolic	V	roots
*ggs*1.3		cytosolic	VI	roots
*ggs*1.4		cytosolic	VII	roots

This arrangement of genes was confirmed by searching in the inbred line R41 (Domínguez 1996) for sequences homologous to those amplified from cv. Peredovic. Gene-specific probes were prepared from the 3' untranslated regions of the four cDNAs for cytosolic isoforms and hybridized to genomic DNA from R41 under stringent conditions. A clear signal of hybridization was observed for all four types of sequences (Fig. 7*A*).

The high similarities showed by the cDNAs for chloroplastic isoforms along their entire sequences disapproved the preparation of cDNA-specific probes. However, it was possible to design cDNA-specific primers against very short unique regions. When cDNA samples from leaves of R41 were subjected to PCR with specific primers, only the cDNA type II was amplified (Fig. 7B). The absence of the cDNAs III and IV from the inbred line R41 confirmed their allelic character.

The size of the plant GS gene families does not show great variations in spite of the diversity of the species considered. Both *Arabidopsis* and pea contain four GS genes (Coruzzi *et al.* 1991, Peterman and Goodman 1991), five GS genes were described in *Phaseolus* (Forde and Cullimore 1989), and six in maize (Li *et al.* 1993). The sunflower GS gene family is one of the largest described so far, including, at least, five members. If additional GS genes exist in the sunflower genome, they either will be similar enough as not being distinguished from the rest by the restriction

analysis performed, or sufficiently weak in their expression that the amount of cDNA obtained was below the limit of detection of the methodology used. In relation to this, it is expected that the extraordinary sensitivity of PCR (Pérez-Vicente *et al.* 1996) should guarantee an exhaustive retrieval of GS sequences from leaves and roots of sunflower.

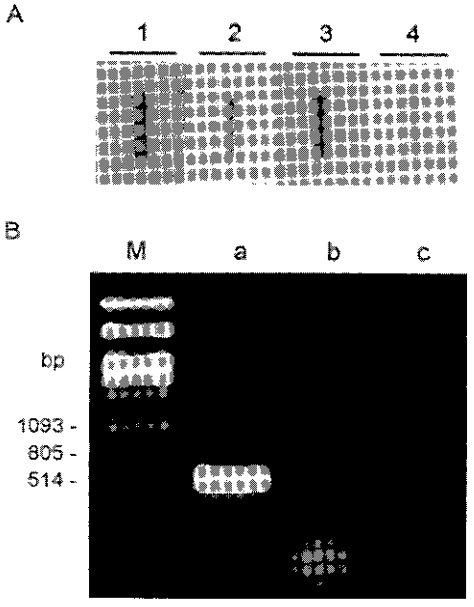

Fig. 7. Survey of GS-coding sequences in the sunflower inbred line R41. *A* - twenty micrograms of genomic DNA from the inbred line R41 were slot blotted onto nylon filters and hybridized with 3' untranslated probes especific for cDNA I (1), cDNA V (2), cDNA VI (3) and cDNA VII (4). *B* - total RNA from leaves of the inbred line R41 was retrotranscribed to cDNA as indicated in Materials and methods. cDNA samples were then amplified by PCR with gs5 and the corresponding specific primer for cDNA II (a), cDNA III (b) or cDNA IV (c). 5 mm^3 from each PCR reaction were loaded per lane. M - molecular size marker corresponding to 0.5 μg of λ phage DNA digested with *Pst* I.

The genes for GS are among the best conserved sequences in plants (Forde and Cullimore 1989). Therefore, the high values of sequence identity obtained in the comparison of the carboxy-terminal sequences of the GS subunits between sunflower and the other plants were not unexpected (Table 3). In general, a higher sequence identity corresponded to a higher phylogenetic proximity, the lowest values being obtained with the far related green algae *Chlamydomonas*. Within the higher plants, dicots always showed higher values than monocots.

According to our data, the cDNA corresponding to *ggs*1.2 was the most divergent of the sequences for cytosolic isoforms (Fig. 5, Table 1). Some of the differences found among the product of *ggs*1.2 and of the others GS genes from sunflower affected amino acids in regions involved in the formation of the active site (Fig. 4). An additional evidence of the uniqueness of the product of *ggs*1.2 was its consistent

divergence from the corresponding sequences of others plants. Compared with the other selected sequences (*ggs*1.1 and *ggs*2a), *ggs*1.2 always gave the lowest values of sequence identity (Table 3).

Table 3. Comparison of the C-terminal end of chloroplastic and cytosolic GS subunits from sunflower to the corresponding sequences from different monocots, dicots and from the green algae *Chlamydomonas reinhardtii*. The amino acid sequences of the coding region of cDNAs I, V and II (genes *ggs*1.1, *ggs*1.2, and *ggs*2a, respectively) were compared to the corresponding sequences of the GS subunits from *Raphanus sativus* (Rap), *Brassica napus* (Bra), *Lycopersicon esculentum* (Lyc), *Nicotiana plumbaginifolia* (Npl), *Pisum sativum* (Pis), *Phaseolus vulgaris* (Pha), *Zea mays* (Zea), *Hordeum vulgare* (Hor), *Arabidopsis thaliana* (Ara), *Nicotiana sylvestris* (Nsy), and *Chlamydomonas reinhardtii* (Chl). The number after the abbreviation indicates cytosolic (1) or chloroplastic (2) subunits. For each GS subunit, the GenBank accesion number is listed in parentheses. The values of sequence identity were calculated as the number of conserved positions divided by the length of the shorter sequence.

Cytosolic GS	*ggs*1.1	*ggs*1.2	Chloroplastic GS	*ggs*2a
Rap1(d25325)	0.94	0.89	Ara2(s69727)	0.94
Bra1(x76736)	0.93	0.88	Bra2(x72751)	0.94
Lyc1(u14754)	0.93	0.87	Pha2(x12738)	0.93
Npl1(m19055)	0.91	0.85	Pis2(m20664)	0.92
Pis1(m20663)	0.91	0.85	Lyc2(u15059)	0.91
Pha1(x04001)	0.91	0.83	Nsy2(x66940)	0.88
Zea1(x65926)	0.90	0.83	Zea2(x65931)	0.91
Hor1(x69087)	0.89	0.85	Hor2(x53580)	0.90
Chl1(u46207)	0.63	0.62	Chl2(u46208)	0.54

In summary, multiple cDNAs derived from a variety of GS genes that are expressed in the leaves and roots of sunflower have been cloned by using a time-saving technique based in the simultaneous amplification of related sequences. This methodology is expected to be of immediate application to new plant species, since the key element, the GS-specific degenerate primer gs5, and others primers of similar characteristic that have proven successful amplifying GS sequences from a wide variety of plants, are available (Pérez-Vicente *et al.* 1996). The impossibility to obtain a full-length cDNA in a single amplification reaction, the major drawback of PCR-cloning, has been palliated by a careful selection of the fragment to be amplified. Although the gs5-QoR fragments amplified from sunflower do not span a full-length cDNA, they contained the information needed to confirm their identity, determine the type of isoform encoded and provide unique regions of sequence to prepare gene-specific probes and primers, thus completing the basic requirements for the characterisation of a multigene family. Once the sequence of the 3' ends is known, the cloning of the corresponding 5' ends can be easily carried out by means of well established methodologies such as RACE (Frohman *et al.* 1988, Frohman 1994, Chenchick *et al.* 1995).

References

Bennett, M.J., Lightfoot, D.A., Cullimore, J.V.: cDNA sequence and differential expression of the gene encoding the glutamine synthetase polypeptide of *Phaseolus vulgaris* L. - Plant mol. Biol. **12**: 553-565, 1989.

Berg, P., Singer, M.: Some genes move around. - In: Berg, P., Singer, M. (ed.): Dealing with Genes. The Language of Heredity. Pp. 161-170. University Science Books, Mill Valley 1992.

Cabello, P., de la Haba, P., Maldonado, J.M.: Isoforms of glutamine synthetase in cotyledons, leaves and roots of sunflower plants. - J. Plant Physiol. **137**: 378-380, 1991.

Chenchick, A., Moqadam, F., Siebert, P.D.: Marathon cDNA amplification: A new method for cloning full length cDNAs. - ClonTechniques **10**(1): 5-8, 1995.

Chou, Q., Russel, M., Birch, D.E., Raymond, J., Bloch, W.: Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. - Nucl. Acids Res. **20**: 1717-1723, 1992.

Coruzzi, G.M., Edwards, J.W., Walker, E.L., Tsai, F-Y., Brears, T.: Regulation of genes for enzymes along a common nitrogen metabolic pathway. - In: Hermann, R.G., Larkins, B.A. (ed.): Plant Molecular Biology. Vol. 2. Pp. 139-145. Plenum Press, New York 1991.

Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H., Bedbrook, H.: mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. - Nucl. Acids Res. **5**: 2229-2240, 1986.

De la Haba, P., Cabello, P., Maldonado, J.M.: Glutamine-synthetase isoforms appearing in sunflower cotyledons during germination. - Planta **186**: 577-581, 1992.

Domínguez, J.: R-41, a sunflower restorer inbred line, carrying two genes for resistance against a highly virulent Spanish population of *Orobanche cernua*. - Plant Breed. **115**: 203-204, 1996.

Edwards, J.W., Coruzzi, G.M.: Photorespiration and light act in concert to regulate the expression of the nuclear gene for the chloroplastic glutamine synthetase. - Plant Cell **1**: 241-248, 1989.

Eisenberg, D., Almassy, R.J., Janson, C.A., Chapman, M.S., Suh, S.W., Cascio, D., Smith, W.W.: Some evolutionary relationships of the primary biological catalyst glutamine synthetase and RuBisCo. - Cold Spring Harbor Symp. Quant. Biol. **52**: 483-490, 1987.

Forde, B.G., Cullimore, J.V.: The molecular biology of glutamine synthetase in higher plants. - Plant mol. Cell. Biol. **6**: 247-296, 1989.

Frohman, M.A., Dush, M,K., Martin, G.R.: Rapid production of full-length cDNAs from rare transcripts by amplification using a single gene-specific oligonucleotide primer. - Proc. nat. Acad. Sci. USA **85**: 8998, 1988.

Frohman, M.A.: Cloning PCR Products. - In: Mullis, K.B., Ferré, F., Gibbs, R.A. (ed.): The Polymerase Chain Reaction. Pp. 14-37. Birkhäuser, Boston 1994.

Hewitt, E.J.: Sand and Water Culture Methods Used in the Study of Plant Nutrition. - Commonwealth Bureau of Horticultural and Plantation Crops, East Malling 1966.

Konat, G.W., Laszkiewicz, I., Grubinska, B., Wiggins, R.C.: Generation of labeled DNA probes by PCR. - In: Griffin, H.G., Griffin, A.M. (ed.): PCR Technology: Current Innovations. Pp. 37-42. CRC Press, Boca Raton 1994.

Lam, H.M., Coschigano, K.T., Oliveira, I.C., Melo-Oliveira, R., Coruzzi, G.M.: The molecular-genetics of nitrogen assimilation into amino acids in higher plants. - Annu. Rev. Plant Physiol. Plant mol. Biol. **47**: 569-93, 1996.

Li, M., Villemur, R., Hussey, P.J., Silflow, C.D., Gantt, J.S., Snustad, D.P.: Differential expression of six glutamine synthetase genes in *Zea mays*. - Plant mol. Biol. **23**: 401-407, 1993.

Li, Q., Hunt, A.G.: The polyadenylation of RNA in plants. - Plant Physiol. **115**: 321-325, 1997.

Miflin, B.J., Lea, P.J.: Ammonia assimilation. - In: Miflin, B.J. (ed.): The Biochemistry of Plants. Pp. 169-202. Academic Press, New York 1980.

Pérez Rodríguez, J., Valpuesta, V.: Expression of glutamine synthetase genes during natural senescence of tomato leaves. - Physiol. Plant. **97**: 576-582, 1996.

Pérez-Vicente, R., Dorado, G., Maldonado, J.M.: Cross-species amplification of glutamine synthetase cDNA by polymerase chain reaction with degenerate primers. - Physiol. Plant. **98**: 705-713, 1996.

Peterman, T.K., Goodman, H.M.: The glutamine synthetase gene family of *Arabidopsis thaliana*: light-regulation and differential expression in leaves, roots, and seeds. - Mol. gen. Genet. **230**: 145-154, 1991.

Roche, D., Temple, S.J., Segupta-Gopalan, C.: Two classes of differentially regulated glutamine synthetase genes are expressed in the soybean nodule: a nodule-specific and a constitutively expressed class. - Plant mol. Biol. **22**: 971-983, 1993.

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

Sukanya, R., Li, M-G., Snustad, D.P.: Root- and shoot-specific responses of individual glutamine synthetase genes of maize to nitrate and ammonia. - Plant mol. Biol. **26**: 1935-1946, 1994.

Temple, S.J., Heard, J., Ganter, J., Dunn, G., Segupta-Gopalan, C.: Characterization of a nodule-enhanced glutamine synthetase from alfalfa: nucleotide sequence, *in situ* localization, and transcript analysis. - Mol. Plant-Microbe Interact. **8**: 218-227, 1995.

Walker, E.L., Coruzzi, G.M.: Developmentally regulated expression of the gene family for cytosolic glutamine synthetase in *Pisum sativum*. - Plant Physiol. **91**: 702-708, 1989.

Wallsgrove, R.M., Turner, J.C., Hall, N.P., Kendall, A.C., Bright, S.W.J.: Barley mutants lacking chloroplast glutamine synthetase - biochemical and genetic analysis. - Plant Physiol. **83**: 155-158, 1987.

Watanabe, A., Kazuyuki, H., Hiroyuki, Y., Watanabe, A.: Biphasic and diferential expression of cytosolic glutamine synthetase genes of radish during seed germination and senescence of cotiledons. - Plant mol. Biol. **26**: 1807-1817, 1994.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A., Weiner, A.M.: Recombination at the molecular level. - In: Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A., Weiner, A.M. (ed.): Molecular Biology of the Gene. Pp. 313-339. The Benjamin Cummings Publ. Comp., Menlo Park 1987.

Yang, J.L., Maher, V.M., McCormick, J.J.: Amplification and direct nucleotide sequencing of cDNA from the lysate of low numbers of diploid human cells. - Gene **83**: 347-354, 1989.