Journal of Plant Physiology 167 (2010) 1101–1108

Contents lists available at ScienceDirect

Journal of Plant Physiology

journal homepage: www.elsevier.de/jplph

A distinct subfamily of papain-like cystein proteinases regulated by senescence and stresses in *Glycine max*

Belén Esteban-García, José Antonio Garrido-Cárdenas, Diego López Alonso, Federico García-Maroto*

Grupo de Biotecnología de Productos Naturales (BIO-279), Facultad de Ciencias Experimentales, Universidad de Almería, 04120 Almería, Spain

ARTICLE INFO

Article history:
Received 22 January 2010
Received in revised form 3 March 2010
Accepted 3 March 2010

Keywords:
Nonsense-mediated mRNA decay
Papain-like protease
Plant stress
Senescence
Soybean

ABSTRACT

GMCP3 encodes a cystein proteinase of *Glycine max* belonging to the papain-like family (C1A in MEROPS database) that was previously found to be involved in the mobilization of protein reserves during seed germination. Here, we report that *GMCP3* is induced by senescence and diverse stresses in non-seed tissues, thus indicating a more general function in plants. Cladistic analysis of papain-like proteins of plants indicated that GMCP3, along with related proteases of other species, belongs to a distinct new group within the C1A family, which can also be distinguished by the four-exon structure of the gene. We also describe the genomic organization of *GMCP3* revealing the presence of two closely related copies that are transcriptionally regulated in a similar way, although only one appears to be functional.

Introduction

Proteinase enzymes catalyze either limited site-specific or complete hydrolysis on their substrates, acting in coordination with other post-translational modifications to regulate turnover, direction and activity of proteins. Among these, cysteine (sulfhydryl) proteinases (CPs) are currently receiving much attention due to their involvement in very diverse aspects of plant physiology and development. They play a role in protein storage and mobilization, senescence, programmed cell death, and are regulated by various types of environmental stresses (revised in Grudkowska and Zagdanska, 2004; Schaller, 2004). Involvement of proteolytic enzymes, including CPs, in the ubiquitin (Ub)/26S proteasome proteolytic pathway is shedding light on the connection of these components to many aspects of plant biology including senescence, cell-cycle, embryogenesis, flower morphogenesis, hormone signaling, and defense (Avci et al., 2008; Vierstra, 2009).

The most investigated CPs are members of the papain-like group (C1A subfamily in the MEROPS catalogue http://merops.sanger.ac.uk). Along with legumains (C13 family), papain-like enzymes are the major CP in plants. C1A proteases are synthesized as relatively inactive precursors (40–50 kDa) containing an N-terminal signal peptide responsible for endoplasmic reticulum (ER) translocation, a variable pro-sequence with regulatory functions in folding and activity, and a 22–35 kDa mature enzyme. Protease activation takes place through limited intra- or inter-molecular proteolysis (Wiederanders, 2003). The structure of proteins from the C1 family exhibits the typical papain-like fold (Drenth et al., 1968), composed of two domains, an α-helix-rich (L) domain and a β-barrel-like (R) domain, separated by a groove containing the catalytic dyad formed by residues Cys25 and His159 (papain numbering), each one on each domain. Two other residues also seem to be important Gln19 aids the formation of the 'oxyanion hole' and Asn175 is important for proper orientation of the His imidazolium ring (Beers et al., 2004).

A number of genes encoding C1A peptidases have been identified in diverse plant species. In *Arabidopsis thaliana*, for instance, the papain-like family comprises 30 genes that have been classified into eight distinct groups (C1A-1 to C1A-8) according to their gene structures and phylogenetic relationships (Beers et al., 2004). Although information about the roles of individual genes is still fragmentary in this plant, the available data indicate that they are regulated by senescence (SAG2 and SAG12, Hensel et al., 1993; Gan and Amasino, 1995), both drought stress and senescence (RD19 and RD21, Koizumi et al., 1993), programmed cell death (CEP genes, Gietl and Schmid, 2001), and during xylogenesis (XCP genes, Funk et al., 2002). More recently, involvement of RD19 in pathogen resistance response has been demonstrated (Bernoux et al., 2008).

GMCP3 is a papain-like protease previously identified in *Glycine max* (Nong et al., 1995). Preliminary gene expression studies showed that the *GMCP3* transcript is stored in the dry seeds, and fur-

Abbreviations: 4-MU, 4-methylumbelliferone; ABA, abscisic acid; CP, cystein proteinase; CSPD, Disodium 3-(4-methoxyspiro {1,2-dioxetane-3,2′-(5′-chloro)tricyclo [3.3.1.13,7]decan}-4-yl)phenyl phosphate; DIG, digoxigenin; ER, endoplasmic reticulum; EtBr, ethidium bromide; IPCR, inverse PCR; MeJa, methyl-jasmonate; NMD, nonsense-mediated mRNA decay; nts, nucleotides; PTC, premature termination codon.

* Corresponding author. Tel.: +34 950015033; fax: +34 950015008.
E-mail address: fgmaroto@ual.es (F. García-Maroto).

0176-1617/$ – see front matter © 2010 Elsevier GmbH. All rights reserved.
doi:10.1016/j.jplph.2010.03.012

ther degraded very early during germination. A possible role in the initiation of storage protein mobilization has been suggested (Nong et al., 1995). Here, we report further characterization of *GMCP3*, including data about the genomic organization and gene structure, as well as relationships to other papain-like proteases. Expression analysis results also showed regulation of *GMCP3* by senescence and different stresses, thus indicating a more general role in the plant.

Materials and methods

Biological material and plant treatments

The Osumi variety of *Glycine max* was used in all the experiments. Seedlings (2–4 true leaves stage) were grown at 25 °C under the controlled conditions of culture cabinets with a 16 h light/8 h dark photoperiod and 70% relative humidity. Leaf material from etiolated seedlings was used as a source of genomic DNA. The different organs analyzed were sampled from mature plants cultivated under greenhouse conditions. The plant treatments were performed using plantlets grown in soil up to the stage of two to four developed true leaves. For dehydration experiments, plantlets were laid on dry filter paper for the indicated times. Wounding was performed using a dented forceps applied to the main veins of the leaves. Treatments with abscisic acid (ABA) and methyl-jasmonate (MeJa) (both from Sigma–Aldrich) were accomplished by pulverization of a 100 μM solution in water. At the end of each treatment, leaves were collected, pooling three plants for each experimental point, frozen in liquid N_2 and stored at −70 °C until RNA extraction.

Experiments performed with transgenic tobacco plants containing the *GMCP3Ψ* promoter fused to the *GUS* gene were performed with plants developed to 8–10 visible leaves stage, except for the assay of senescent leaves, which was carried out on adult plants. Leaf discs, pooled from three plants, were collected and analyzed for GUS activity as indicated below. Stress treatments were performed as in soybean. The results shown correspond to a representative transgenic line, but the same regulation was confirmed in at least two other lines, depending on the treatment.

Cloning of the GMCP3 and GMCP3Ψ genes

The genomic sequence comprising the whole coding region of *GMCP3* was assembled from two overlapping clones obtained by PCR on genomic DNA. The two DNA fragments of 1.5 kb and 1.3 kb were generated using the oligonucleotide pairs OCEP-1 (5′-ATGGAGGCAAAGCGAGGCCATGCCCTCAT-3′)/OCEP-2 (5′-TTGTTGTCACAGTCAAGGAGCTGTTGTTCAC-3′), and OCEP-3 (5′-GTGCGAGATAGGTAGAAGGAACAGAGTG-3′)/OCEP-4 (5′-CAGAACTAAGGAAAACCAAAGAGGAGATGC-3′), respectively. PCR amplification was conducted using a proofreading polymerase (Pfx®, Invitrogen) and a program consisting of a denaturation step of 2 min at 94 °C, followed by 35 cycles of 15 s at 94 °C, 45 s at 60 °C and 2 min at 68 °C, ending with a 7 min step at 68 °C. The resulting fragments were cloned in the vector pGEM-T-Easy® (Promega) and several clones were fully sequenced. A full cDNA clone was also obtained for *GMCP3*. Briefly, a cDNA was synthesized, as described below, using total RNA obtained from developing soybean seeds, and RT-PCR amplification was performed using the flanking oligonucleotide primers OCEP-1 and OCEP-4. The sequence of the 3′-untranslated region was completed by 3′-RACE as indicated below.

Cloning of *GMCP3Ψ* was accomplished by obtaining of a 1.8 kb partial fragment by genomic PCR using primers OCEP-5 (5′-CAAGCAAGCACAGAGATGGAGGCAAAG-3′) and OCEP-6 (5′-GACCTCCATTACACCCATTGTCACATG-3′), followed by bi-directional genomic walking through inverse PCR (IPCR). The method by Ochman et al. (1990) was employed with minor modifications (Garcia-Maroto et al., 2007). Thus, from the partial sequence of *GMCP3Ψ*, two nested upstream primers OCEP-7 (5′-GAGCTAAGCACGTGATGGAATGGTCTC-3′) and OCEP-8 (5′-AGCAGAGGATAGGGTTAGAGCGAAGAG-3′), and two nested downstream primers OCEP-9 (5′-TCTGCGCACGGAGAAGAAGTTCAAGGT-3′) and OCEP-10 (5′-CATGGAGAATTACGGGAGGAGCTACTC-3′), were designed to perform the IPCR. Either EcoRI or VspI restriction enzymes were used to cut genomic DNA. Released fragments were circularized with T4 DNA ligase and subjected to two nested rounds of PCR amplification, allowing us to obtain 1.6 kb (from EcoRI digestion) and 2.5 kb (VspI digestion) -long products that were useful for the walking process. These fragments were analyzed and sequenced as described above, allowing the assemblage of the whole coding sequence for *GMCP3Ψ*, including 1.4 kb of the 5′-flanking region.

The sequences have been deposited in the GenBank under the accession numbers GU452503 (*GMCP3* cDNA), GU452501 (*GMCP3* gene) and GU452502 (*GMCP3Ψ* gene).

RT-PCR and 3′-RACE analysis

The total RNA, employed for RT-PCR and RACE analysis, was purified from different soybean tissues (leaves and developing seeds) using the RNeasy Plant Mini Kit (QUIAGEN). Elimination of contaminating genomic DNA was performed using DNase-I (RQ1 RNase-Free DNase, Promega). The cDNA for RT-PCR was synthesized from 1 to 8 μg of RNA using oligo(dT) and employing the SuperScript First-Strand Synthesis System for RT-PCR kit (Invitrogen), following manufacturer's instructions.

Semi-quantitative RT-PCR was performed using the oligonucleotide primers UP-N (5′-ATGGAGGCAAAGCGAGGCCATGCCCTCAT-3′) and UP-Ψ (5′-ATGGAGGCAAAGCGAGACCATTCCATCAC-3′) as specific upstream primers for *GMCP3* and *GMCP3Ψ*, respectively, located on the first exon, in combination with the C13 primer (5′-CTTAACTTCGGTGACAGCTCCTTTCTCTC-3′) common to both genes and located on the second exon (see Fig. 4A). The PCR reactions were performed in 25 μL containing cDNA derived from 0.5 μg of RNA from developing seeds, using a hot-start Taq polymerase (Taq Platinum®, Invitrogen) and a program consisting of a denaturation step of 2 min at 94 °C, followed by 25–36 cycles of 15 s at 94 °C, 30 s at 58 °C and 1 min at 72 °C, ending with a 5 min step at 72 °C. Identity of the respective amplification products was confirmed by direct sequencing of the fragments.

Similarly, relative expression of *GMPCP3* and *GMCP3Ψ* in plantlets subjected to dehydration (experiment in Fig. 5C) was estimated by RT-PCR using either UP-N or UP-Ψ as the upstream specific primers in combination with C11 primer (5′-GAGAGAAAGGAGCTGTCACCGAAGTTAAG-3′), common to both genes and located at the end of the first exon. PCR conditions were as described above, using 32 cycles. As a loading control, amplification of the actin-2 was carried out in parallel on equivalent samples using the oligonucleotide primers ACT2-F (5′-TCCCTCAATCTCATCTTCTTCC-3′) and ACT2-R (5′-GACCTGCCTCATCAATCTTCTTCC-3′).

3′-RACE was performed using the SMART RACE® kit (Clontech), following manufacturer's protocols. 1–2 μg of total RNA from developing seeds (Fig. 4C and Suppl. Fig. 2), or leaves from normal and dehydrated seedlings (Suppl. Fig. 2) were used for the synthesis of the oligo(dT) primed cDNA linked to an appropriate adaptor. Either the UP-N primer specific for *GMCP3*, UP-Ψ and C17 (5′-GAGAAGCTCTACACCGGCTCCGTCGAC-3′) for *GMCP3Ψ*, or the C6 (5′-TCTGCGCACGGAGAAGAAGTTCAAGGT-3′) primer, common to both *GMCP3* and *GMCP3Ψ*, were used as sense gene-specific primers for PCR amplification, as indicated in the experiments. The

PCR program comprised 30 cycles of 30 s at 94 °C, 30 s at 68 °C and 3 min at 72 °C.

RNA and DNA gel blot analysis

Total RNA was extracted from different soybean tissues using the Concert® Plant RNA Reagent (Invitrogen) following the protocol provided by the manufacturer. About 10 μg per lane of total RNA were loaded onto an agarose/formaldehyde gel, separated electrophoretically, and transferred to Hybond®-N+ membranes. Filters were UV-crosslinked, pre-hybridized at 50 °C for 5 h in the 50% formamide/high SDS buffer recommended by the digoxigenin (DIG) manufacturer (Roche), and hybridized at the same temperature and same buffer solution (stringent conditions). A DIG-labeled GMCP3-specific probe was generated by random primed labeling of a 300 bp fragment corresponding to the 5′-end of the GMCP3 coding sequence. High-stringency washes were performed twice at 68 °C for 15 min in buffer containing 0.1× SSC, 0.1% SDS, and the luminogenic substrate CSPD® was used for the detection, following the instructions provided with the DIG detection kit. Images were registered using a Chemie Genius2 detection system (SynGene).

DNA gel blot of the PCR products obtained in the 3′-RACE analysis was conducted using the same DIG-labeled probe as above and following a similar protocol. In this case, pre-hybridization and hybridization were performed in 50% formamide/high SDS buffer at 42 °C, and washes were performed at 65 °C.

Cladistic analysis

A search for CPs of the C1A family (MEROPS catalog) was performed in GenBank using the CP of *Arabidopsis* representative of the different subfamilies previously defined (Beers et al., 2004) as inputs for independent searches. A total of 165 CPs from dicotyledonous species were gathered and further aligned, using the program Clustal X v.1.7 (Thompson et al., 1997) under the default settings. The alignment output was used to generate an unrooted phylogenetic tree based on the Minimum Evolution method (Rzhetsky and Nei, 1992), as implemented in the MEGA package v4.0 (Tamura et al., 2007). The Poisson correction metric was used together with the pairwise deletion option, and confidence of the tree branches was checked by bootstrap generated from 1000 replicates.

Transformation of tobacco with GMCP3Ψ promoter fused to GUS

The sequence, about 1.4 kb, of the 5′-flanking region of *GMCP3Ψ* was obtained by IPCR as indicated above. To isolate this region, a PCR was performed and the resulting fragment was cloned in the pBluescript (KSII) plasmid and sequenced. A SalI–BamHI fragment was excised from KSII and cloned into the pCAMBIA 1391Z binary vector (CAMBIA, Australia). The T-DNA region of this plasmid contains a promoterless *GUS* gene interrupted by a catalase intron. The resulting construct was used to transform the *A. tumefaciens* LBA4404 strain. Tobacco leaf disc transformation was achieved essentially as described by Horsch et al. (1985). Homozygous lines containing a single T-DNA insertion were selected for the experiments based on segregation of the hygromycin resistance in the progeny.

Fluorometric determination of GUS activity

Leaf discs were collected and ground in Eppendorf tubes with a pestle and sand. After grinding, 0.5 mL of extraction buffer (125 mM KH_2PO_4/K_2HPO_4 pH 7.0, 100 mM Na_2EDTA, 1 mM DTT, 0.5% v/v Triton X-100, 0.5% w/v Laurylsarcosine) was added and centrifuged for 10 min (13,000 rpm) at 4 °C. A 5 μL aliquot of the supernatant was completed with 20 μL of extraction buffer and mixed with 25 μL of pre-warmed (37 °C) GUS assay solution (2 mM methyl-4-umbelliferyl-β-D-glucuronide in extraction buffer). The mixture was incubated at 37 °C for 1 h, and 15 μL were withdrawn and added onto 1 mL of stop buffer (0.2 M Na_2CO_3). Stop buffer and 10–100 nM 4-methylumbelliferone (4-MU) were used for calibration and standardization. The fluorescence was determined using a microtiter plate reader fluorometer (Mod. FLx800, BIO-TEK Instruments Inc) at 360 and 465 nm as excitation and emission filter wavelengths, respectively. The protein concentration of the extracts was estimated by the Lowry method (Lowry et al., 1951). GUS-specific activity in the experiments was expressed as nmol of released 4-MU per min and mg of protein extract under our standard conditions.

Results

Structure and genomic organization of the GMCP3 gene

The sequence of the *GMPC3* gene, spanning about 2.5 kb of genomic DNA, was obtained as described in 'Materials and methods' section. The nucleotide sequence deduced for the coding region was identical to that of the cDNA (Acc. No. GU452503). Gene structure, obtained by comparison to cDNA, revealed the presence of four exons (Fig. 1A). Sequences around intron/exon boundaries were found to conform the usual consensus splice sites, and are in good agreement to those predicted by software tools such as GENSCAN (Burge and Karlin, 1997).

GMPC3 encodes a protein of 380 residues, with a predicted molecular mass of 41.6 kDa. Both the signal peptide (residues 1–29) and propeptide (30–139) were encoded within the first exon of the gene, while the mature protein was contained mainly (except for the first 19 residues) in exons 2–4 (Fig. 1A).

In the course of cloning *GMCP3*, a closely related gene was also detected and was studied further. The genomic sequence obtained for this gene (see 'Materials and methods' section) shares 95% of identical nucleotides of the putative coding sequence with that of *GMCP3*. According to the existence of two closely related *GMCP3* genes, DNA gel blot analysis revealed the presence of multiple hybridization bands under stringent conditions (results not shown), as also shown in a previous report (Nong et al., 1995). A detailed characterization of the *GMCP3*-related gene, hereafter designated *GMCP3Ψ*, revealed a 17 nts deletion at the end of the first exon (Fig. 1B). This generates a premature termination codon (PTC) placed 358 nts downstream the ATG, and 100 nts from the predicted splicing site located downstream (Fig. 1A).

GMCP3 defines a distinct subgroup within the C1A family of proteases

GMCP3 encodes a cistein proteinase belonging to the C1A family (MEROPS catalogue), which groups papain-like proteases with a catalytic dyad composed of Cys134 and His307 residues (Beers et al., 2004). A database search showed similarity of GMCP3 to CPs CPR4 of *Vicia sativa* (Acc. No. Z99172) and *Phaseolus vulgaris* (Acc. No. Z99955), which are involved in seed development and globulin mobilization during early germination stages (Fischer et al., 2000; Muntz et al., 2001; Muntz and Shutov, 2002). To further analyze relationships of GMCP3 with other CPs, we performed an extensive cladistic analysis including 165 plant proteinases available in GenBank and belonging to the C1A family. The unrooted Minimum Evolution tree from this analysis is depicted in Fig. 2. As shown in this figure, GMCP3 and CPR4 group together with two other uncharacterized proteins from *Vitis vinifera* (CAN64491) and *Arabidopsis thaliana* (AAL49820), making up a clade (99% bootstrap

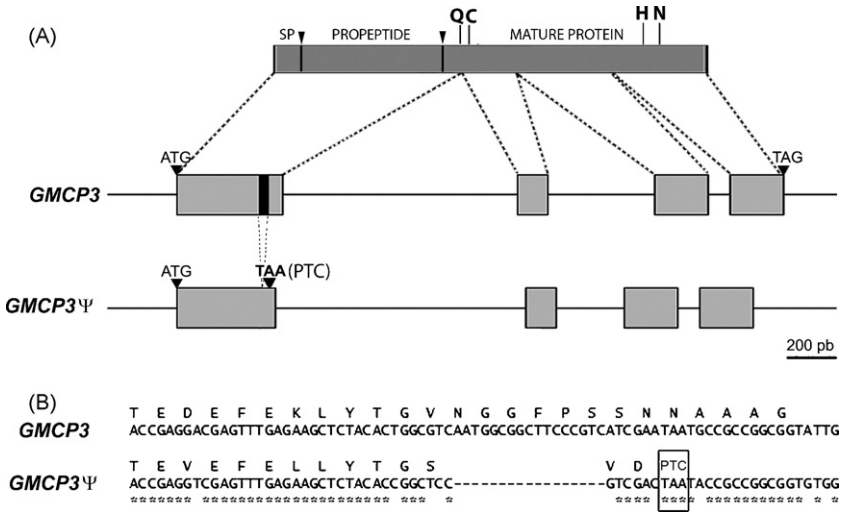

Fig. 1. (A) Deduced genomic structures for the *GMCP3* genes. The position of the PTC within the first exon of *GMCPΨ* is indicated. Correspondence to the protein sequence (dark shadowed), as well as critical Q, C, H and N residues for proteolytic activity (see 'Introduction' section) are also shown. (B) Sequence alignment around the PTC in *GMCP3Ψ* and *GMCP3* genes. The PTC is generated by a 17 nts deletion in *GMCP3Ψ*.

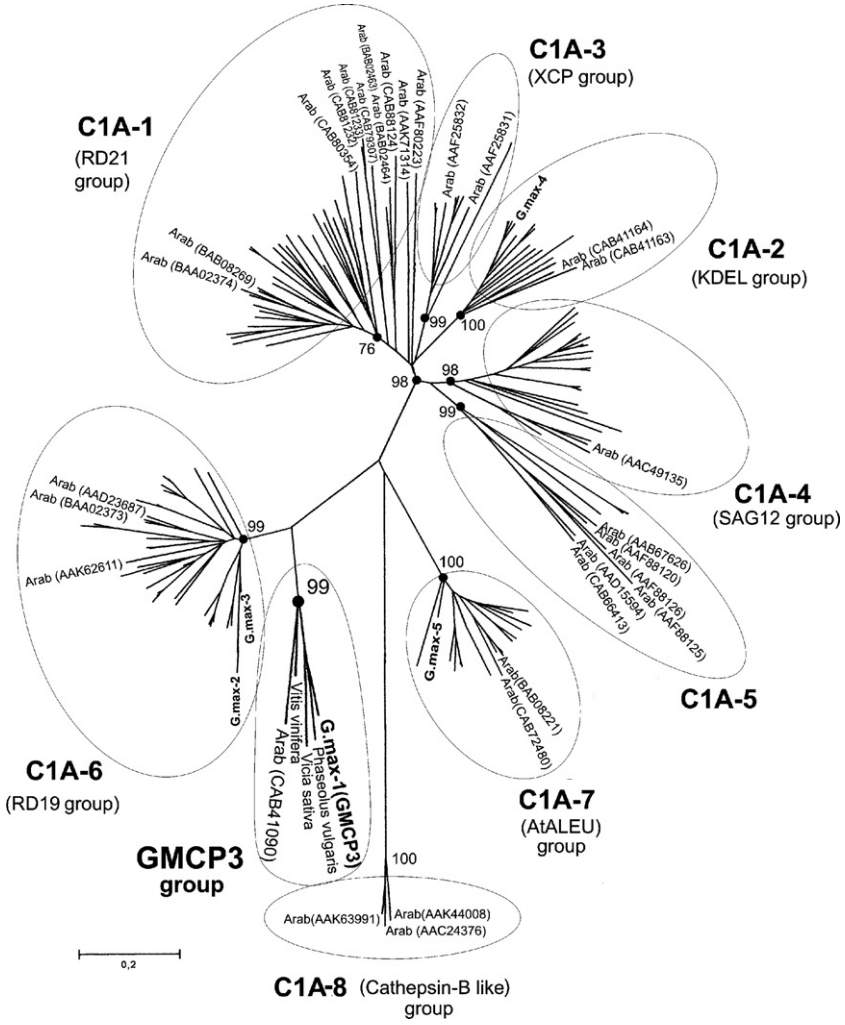

Fig. 2. Unrooted Minimum Evolution cladogram of plant cystein proteinases from the C1A family. Amino acid sequences retrieved from the GenBank were analyzed as described in 'Materials and methods' section, and the different clades designated according to the eight groups (C1A-1 to CA1-8) previously described in *Arabidopsis* (Beers et al., 2004). Sequences from *Arabidopsis* are identified by their accession numbers. Also five different sequences identified in soybean are designated as G. max-1. Putative GMCP3 orthologues within the GMCP3 group are identified by the name of their respective species. Bootstrap values supporting the different groups are indicated on relevant nodes.

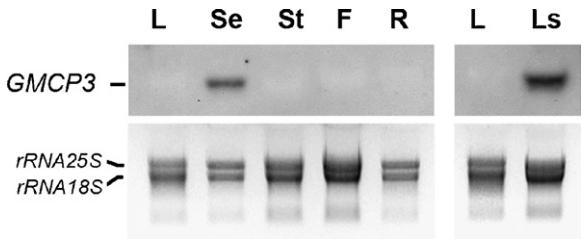

Fig. 3. RNA gel blot analysis of *GMCP3* in different soybean organs. Equivalent amounts (about 10 μg) of total RNA from green leaves (L), developing seeds (Se), stem (St), flowers (F), roots (R) and senescent leaves (Ls), collected from adult plants at anthesis, were subjected to electrophoresis in an agarose/formaldehyde gel and analyzed by RNA gel blot as described in 'Materials and methods' section. Hybridization was performed using a DIG-labeled *GMCP3* probe under high-stringency conditions and the detection with CSPD. The ethidium bromide (EtBr) staining of the gel is shown below as a loading control.

support) clearly differentiated from the eight groups (C1A-1 to C1A-8) previously described in *Arabidopsis* (Beers et al., 2004, Fig. 2). In this tree, the GMCP3 clade appears as a sister group of the C1A-6 subfamily that includes two other papain-like proteins of soybean (designated G. max-2 and G. max-3 in Fig. 2) and the RD19 protein, which is induced by drought in *Arabidopsis* (Koizumi et al., 1993) and is involved in pathogen resistance (Bernoux et al., 2008). Differentiation of the GMCP3 and the C1A-6 groups is also supported by their gene structures. While *GMCP3* and their putative orthologues of *Arabidopsis* (NC_003075), *Vitis* (NC_012009), *Populus* (NC_008475), and *Ricinus* (NW_002994282) share a common gene structure comprised by four exons, members of the CA1-6 group in their respective species have three or four (Beers et al., 2004). Alignment of C1A and GMCP3-related proteins revealed strong similarity between the two groups (Suppl. Fig. 1), with long stretches of conserved residues, including critical residues. However, conspicuous differences were apparent, mainly in a region before the processing site for the propeptide. Thus, a glycine-rich sequence is present at the end of the propeptide in GMCP3-like proteases, while charged residues predominate in the same region of CA1-6 proteins (Suppl. Fig. 1). Divergence within this region might be responsible for differences in the processing of the two groups of CPs.

GMCP3 expression in different organs of the soybean plant

The expression of *GMCP3* in different organs of the soybean plant was investigated by RNA gel blot analysis. As shown in Fig. 3, the *GMCP3* transcript was detected only in developing seeds of normal plants. Interestingly, a high expression level was also observed in senescent leaves, while it was not detectable in green leaves sampled from equivalent positions (same leaf number) of a younger plant (Fig. 3). Since cross-hybridization of the DNA probe was expected from the high similarity of the two related *GMCP3* genes, an evaluation of the individual expression of each gene was carried out using semi-quantitative RT-PCR analysis. When PCR was performed using cDNA from developing seeds together with appropriate specific primer combinations (Fig. 4B), a much lower amount of the amplification product was always obtained for *GMCP3Ψ* as compared to *GMCP3*, a difference that was particularly apparent at lower cycle numbers (Fig. 4B). Similar results were obtained using two other primer combinations (results not shown), indicating that the expression level of *GMCP3Ψ* in the seed tissues was much lower relative to that of the *GMCP3* gene. A low relative expression of *GMCP3Ψ* was also found when RT-PCR was performed on mRNA from senescent leaves (not shown).

To further investigate gene expression and the size of transcripts derived from each gene, 3′-RACE was performed using the seed cDNA (Fig. 4C). Different specific combinations of a gene-specific upstream primer with the common downstream anchor primer

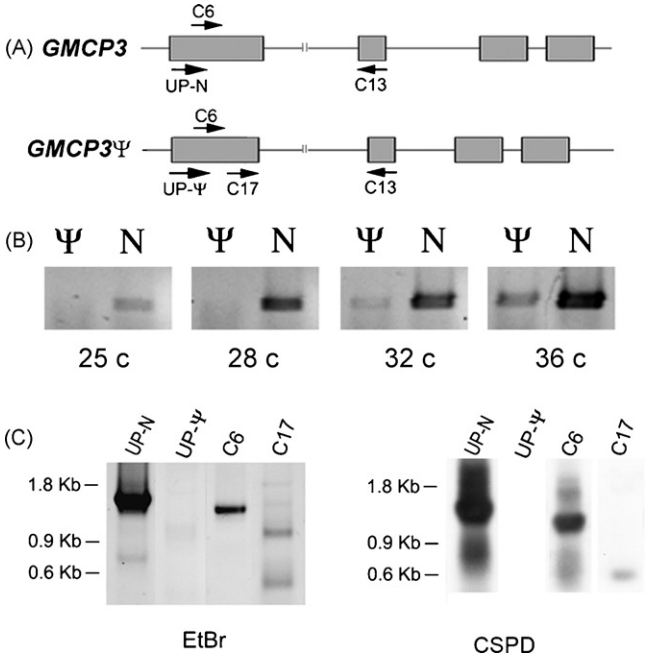

Fig. 4. Expression analysis of *GMCP3* and *GMCP3Ψ* in developing seeds of soybean by semi-quantitative PCR and 3′-RACE. (A) Location scheme of oligonucleotide primers used in the experiments. (B) Semi-quantitative PCR on developing seeds RNA. Equivalent amounts of oligo-dT synthesized cDNA (0.5 μg) were employed for RT-PCR using the UP-N/C13 (N) or UP-Ψ/C13 (Ψ) primer combinations specific for *GMCP3* and *GMCP3Ψ* genes, respectively, for the indicated cycle numbers (25–36 cycles). The identity of the PCR fragments was confirmed by direct sequencing. (C) 3′-RACE analysis on RNA from developing seeds. cDNA obtained using an appropriate oligo-dT anchor primer was subjected to PCR amplification using a nested oligonucleotide in combination with gene-specific primers (UP-N for *GMCP3*, UP-Ψ or C17 for *GMCP3Ψ*, or C6 for both genes; see panel A. Amplification products were separated by electrophoresis and detected either directly by EtBr or by DNA gel blot using a DIG-labeled *GMCP3* probe and development with CSPD.

were assayed (Fig. 4A). Amplification using the *GMCP3*-specific primer UP-N produced a 1.4 kb band of the expected transcript size that hybridized with a *GMCP3* probe. Sequencing of this fragment confirmed the identity of the *GMCP3* transcript, and allowed the identification of a polyadenylation site 256 nts downstream from the stop codon (Acc. No. GU452501). However, no detectable product hybridizing with the *GMCP3* probe was obtained from the same cDNA using either the UP-Ψ or the C17-specific primers for *GMCP3Ψ* (Fig. 4C). Although two bands (of 0.4 and 1 kb) were apparent after PCR with the C17 primer, cloning and sequencing of these fragments showed that they were not related to any of the *GMCP3Ψ* genes. RACE amplification with C16, a common primer for both genes, produced a 1.2 kb band of the expected size, although further analysis of the product revealed that it corresponded exclusively to GMCP3.

The reduced transcript level for *GMCP3Ψ* is in agreement with the presence of a PTC on the first exon of this gene, and indicates that it could be down-regulated through a nonsense-mediated mRNA decay (NMD) mechanism (Hori and Watanabe, 2007).

Regulation of GMCP3 and GMCP3Ψ by plant stresses

We examined the possible regulation of the *GMCP3* gene in soybean seedlings subjected to different stresses. RNA gel blot analysis showed a clear induction of *GMCP3* in plantlets that were dehydrated over 24 or 48 h (Fig. 5A), while expression was not detectable at 12 h or in control plants. Pulverization treatment of plants with 100 μM ABA, a well-known mediator of the drought stress, was also able to trigger a similar response of *GMCP3* (Fig. 5B).

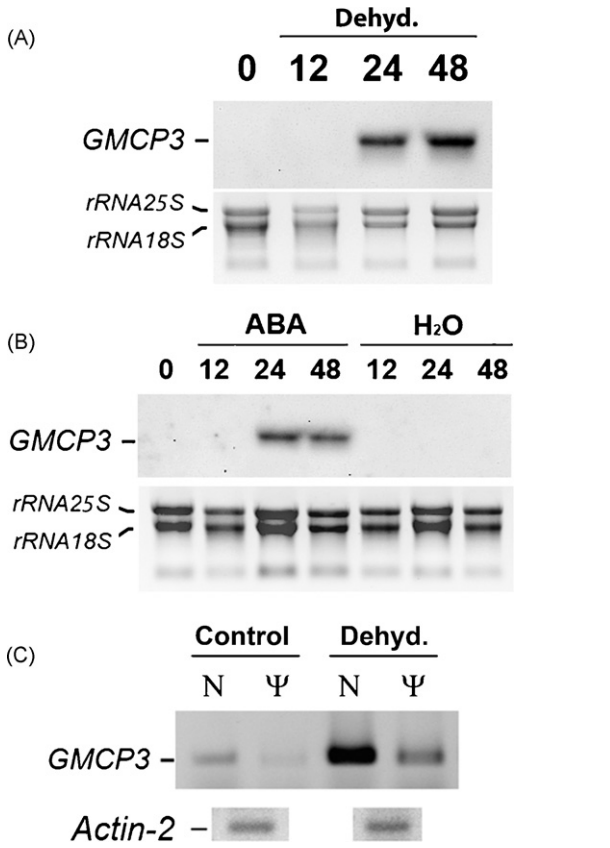

Fig. 5. Expression analysis of *GMCP3* and *GMCP3Ψ* in soybean plantlets subjected to hydric stress or ABA treatment. (A) RNA gel blot analysis of total RNA obtained from plantlets subjected to dehydration for different times (12–48 h). Hybridization was performed under stringent conditions using a DIG-labeled GMCP3 probe, and detection with CSPD. (B) RNA gel blot on total RNA from plantlets collected at different times (12–48 h) after pulverization either with 0.1 mM ABA or with water as a negative control. (C) Semi-quantitative PCR on RNA from control or 48h-dehydrated plantlets. PCR amplification was performed on equivalent amounts of oligo-dT directed cDNA, using the C13 primer in combination with gene-specific primers UP-N (for *GMCP3*, lanes N) or UP-Ψ (for *GMCP3Ψ*, lanes Ψ). As a control of equal template loading amplification with actin-2 primers was performed under the same conditions. PCR products were analyzed by electrophoresis and visualized with EtBr.

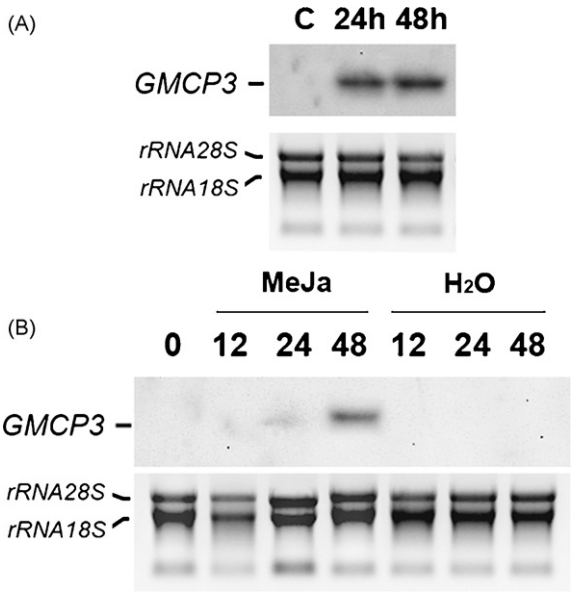

Fig. 6. Expression analysis of *GMCP3* and *GMCP3Ψ* in soybean plantlets subjected to mechanical wounding or MeJa treatment. (A) RNA gel blot analysis of total RNA obtained from plantlets collected at different times (24 or 48 h) after wounding. (B) RNA gel blot on total RNA from plantlets collected at different times (12–48 h) after pulverization either with 0.1 mM MeJa, or with water as a negative control. Hybridization was performed under stringent conditions using a DIG-labeled GMCP3 probe, and the detection with CSPD.

RT-PCR performed with specific primers to individually determine the regulation of *GMCP3* and *GMCP3Ψ* indicated that dehydration increases the expression of both genes, though the transcript level of *GMCP3* under stress conditions was again higher than the levels attained by *GMCP3Ψ* (Fig. 5C). 3′-RACE experiments using specific primers were in agreement with these results and, in addition, they showed that transcripts generated upon stress induction of *GMCP3* were the same size as those present in normal developing seeds (Suppl. Fig. 2).

Regulation of *GMCP3* by wounding and related regulatory molecules was also investigated. As shown in Fig. 6A, induction of *GMCP3* was detected by RNA gel blot analysis from 24 h after leaf wounding. Treatment of plants with MeJa, a key mediator in the octadecanoid pathway that is activated by wounding, also increased the expression of *GMCP3* (Fig. 6B).

Results from RT-PCR experiments indicated that *GMCP3Ψ* might be up-regulated under stress conditions. To further confirm this hypothesis, we fused the 5′-regulatory region of *GMCP3Ψ* (about 1.4 kbp) to the *GUS* reporter gene in the pCAMBIA-1391Z vector, and the *prGMCP3Ψ::GUS* construct was employed to transform tobacco plants via *Agrobacterium tumefaciens*. Experiments carried out with different primary transgenic lines indicated tissue regulation of *GMCP3Ψ* similar to that found for *GMCP3* in soybean, with low GUS activity in vegetative organs relative to the attained level in developing seeds (results not shown). As shown in Fig. 7, a progressive increase of GUS activity was observed along leaf senescence in adult plants (Fig. 7A). Moreover, a clear induction of GUS activity was registered when *prGMCP3Ψ::GUS* plantlets were subjected to wounding (Fig. 7B) or drought stress (Fig. 7C). Similarly, pulverization with MeJa and ABA regulators induced GUS activity (Fig. 7D).

Discussion

In a previous study, a cDNA encoding a papain-like protease in soybean, called GMCP3, was reported (Nong et al., 1995). The initial characterization of *GMCP3* indicated seed-specific expression and a possible role during protein storage mobilization. Data presented here suggest a more general function. We have shown that *GMCP3* is also highly expressed during leaf senescence. Regulation by senescence has been also reported for CPs from diverse subfamilies of papain-like proteinases (Fig. 2), including, for example, SAG2, SAG12, RD19 and RD21 in *Arabidopsis* (Hensel et al., 1993; Koizumi et al., 1993; Gan and Amasino, 1995). These proteases have been classically recognized as effectors in protein remobilization, allowing recycling to other parts of the plant (Smart, 1994; Buchanan-Wollaston, 1997), although a regulatory role in the programmed cell death established during senescence has been more recently proposed (revised in Schaller, 2004). *GMCP3* is also induced by drought stress, similarly to what has been reported for other CPs including RD19 and RD21 (Koizumi et al., 1993). Accordingly, up-regulation by ABA, a key hormonal regulator of the hydric stress pathway, was also observed for *GMCP3*. Although the protein substrates of CPs induced by water deficiency have not been defined, their involvement in rebuilding of damaged or misfolded proteins is generally assumed. We have shown that GMCP3 is also induced by mechanical wounding and the regulator of the octadecanoic pathway MeJa, thus linking its expression to the plant defensive response. Evidence is emerging that papain-like CPs play crucial

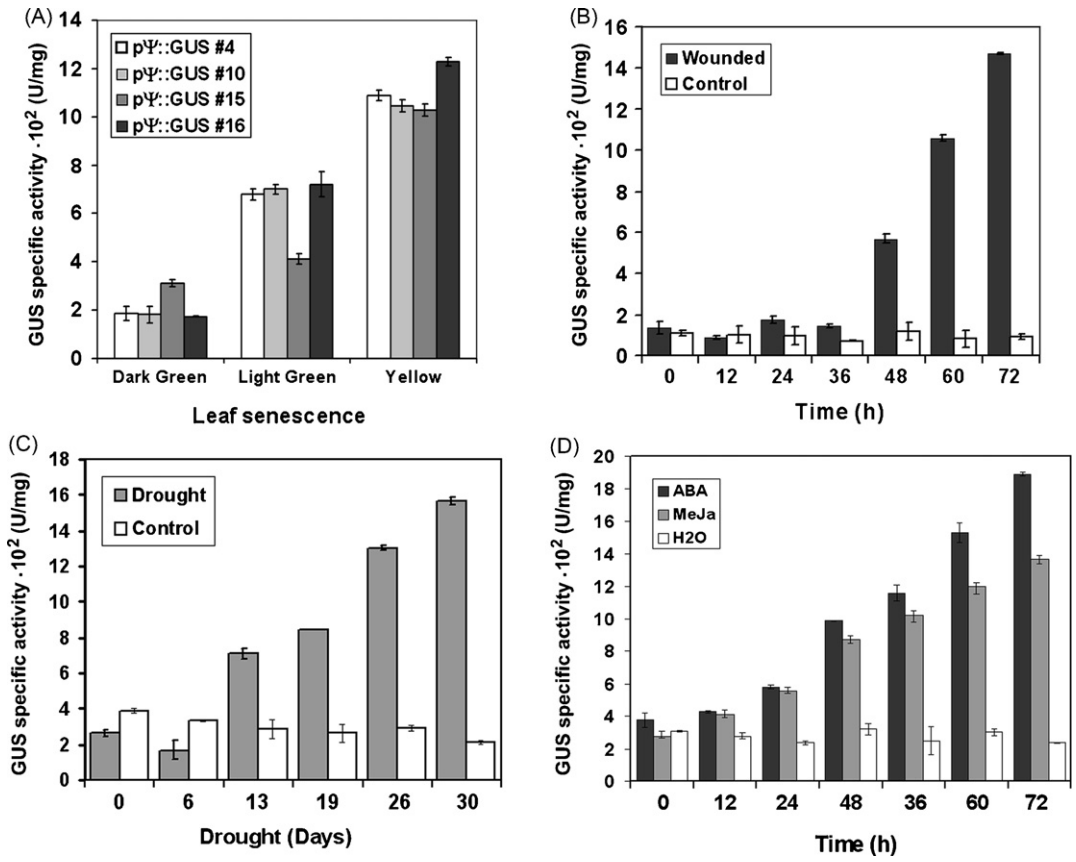

Fig. 7. GUS activity in *GMCP3Ψ::GUS* transgenic tobacco plants under different conditions: (A) leaves from four adult primary transformant plants collected at different senescence states (dark green non-senescent, light green starting senescence, or fully senescent yellow leaves); (B) plants collected at different times (12–72 h) after mechanical wounding. Unwounded plants were taken at the same time as a control. (C) Plants kept without watering for 6–30 days. Watered plants grown in the same conditions were used as a control; (D) plants pulverized with 0.1 mM ABA, MeJA, or just water as a control. GUS measurements were performed on protein extracts as described in 'Materials and methods' section, and the values expressed as specific activity (4-MU nmol min^{-1} mg^{-1}).

roles in plant–pathogen/pest interactions. Some examples are the papain from papaya and Mir1 (maize inbred resistance 1) from maize, both extracellular secreted proteins acting against insect larvae (see Shindo and Van Der Hoorn, 2008 for a review). It has recently been reported that RD19, a CP from the sister clade to the GMCP3 group, is required for the R-protein-mediated resistance to the bacteria *Ralstonia solanacearum* in *Arabidopsis* (Bernoux et al., 2008).

We cloned the genomic sequence of *GMCP3*, revealing a gene structure composed of four exons. Comparison of GMCP3 to other plant CPs allowed the detection of highly related sequences in *Arabidopsis*, *Vitis*, *Populus* and *Ricinus* (Suppl. Fig. 1) making up a well-supported separate cluster. All of these genes shares the same structure, and they likely represent *GMCP3* orthologues. This notion is reinforced by the existence, in the same organisms (also in soybean), of related CPs within the C1A-6 sister clade, and suggests an early divergence of these two lineages in dicots. Previous analysis allowed the identification of eight groups within the C1A family based on sequence similarity and gene structure (Beers et al., 2004). Our cladistic analysis indicated that GMCP3-related proteins make up a separate group within the C1A family for which little information is available from any of these organisms. In *Arabidopsis*, for instance, a cDNA (CAB41090) was annotated encoding a possible orthologue of GMCP3, but it was assumed to be seed-specific (Gruis et al., 2002). However, a more detailed analysis, using the Genevestigator software (Hruz et al., 2008) on available transcriptional profiling data for the *Arabidopsis* gene, indicated similar regulation to *GMCP3* by diverse plant stresses (including drought and wounding) as well as related hormonal signals (ABA and MeJa) (Suppl. Fig. 3), in addition to seed-specific expression under normal conditions (Suppl. Fig. 4).

A search for cis-acting DNA motifs in promoters of the *GMCP3* group is also in agreement with a similar regulation for these genes. Thus, element complexes containing RY and G boxes, which are involved in seed-specific expression and ABA response (Ezcurra et al., 1999), are present within the proximal 5′-regulatory region of the genes (Suppl. Table 1). In addition to G elements, multiple W-boxes classically involved in defense-related processes (Chakravarthy et al., 2003) might be related to the wounding/MeJa response.

We found a closely related copy (*GMCP3Ψ*) of the *GMCP3* gene in soybean. The presence of a PTC at the end of the first exon, in addition to a reduced level of the *GMCP3Ψ* transcript, suggests that this gene represents a 'non-functional' copy (pseudogene) of *GMCP3*. Analysis of the 5′-regulatory region of *GMCP3Ψ* showed that it is able to direct high expression of a reporter gene with a similar regulation pattern to that of *GMCP3*. This indicates some kind of post-transcriptional regulation. It has been found that genes containing a PTC located more than 50 pb of the next intron–exon boundary are subject to down-regulation through a NMD mechanism (Hori and Watanabe, 2007; Wu et al., 2007). Since this is the case for *GMCP3Ψ*, it is likely that reduced expression of the pseudogene is caused by selective degradation of its transcript *via* NMD.

However, a possible role for *GMCP3Ψ* cannot be excluded, even though its transcript level is largely reduced compared to *GMCP3*. It is important to note that eventual translation of the *GMCP3Ψ* mRNA would result in a truncated protein containing the whole propeptide. Given the inhibitory activity described for this domain

(Taylor et al., 1995; Chagas et al., 1996), it is tempting to speculate a possible function of the *GMCP3* paralogue to trans-regulate the protease activity of GMCP3.

Acknowledgements

This work was supported by grants from the Ministerio de Ciencia y Tecnología (MCYT, AGL2005-01498/AGR) and Junta de Andalucía (P05-189-AGR). Belén Esteban-García was recipient of a postgraduate fellowship from the Junta de Andalucía.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jplph.2010.03.012.

References

Avci U, Petzold HE, Ismail IO, Beers EP, Haigler CH. Cysteine proteases XCP1 and XCP2 aid micro-autolysis within the intact central vacuole during xylogenesis in *Arabidopsis* roots. Plant J 2008;56:303–15.

Beers EP, Jones AM, Dickerman AW. The S8 serine, C1A cysteine and A1 aspartic protease families in Arabidopsis. Phytochemistry 2004;65:43–58.

Bernoux M, Timmers T, Jauneau A, Briere C, de Wit PJ, Marco Y, et al. RD19, an *Arabidopsis* cysteine protease required for RRS1-R-mediated resistance, is relocalized to the nucleus by the *Ralstonia solanacearum* PopP2 effector. Plant Cell 2008;20:2252–64.

Buchanan-Wollaston V. The molecular biology of leaf senescence. J Exp Bot 1997;48:181–99.

Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Biol 1997;268:78–94.

Chagas JR, Ferrerdimartino M, Gauthier F, Lalmanch G. Inhibition of cathepsin B by its propeptide: use of overlapping peptides to identify a critical segment. FEBS Lett 1996;392:233–6.

Chakravarthy S, Tuori RP, DAscenzo MD, Fobert PR, Despres C, Martin GB. The tomato transcription factor Pti4 regulates defence-related gene expression via GCC box and non-GCC box cis elements. Plant Cell 2003;15:3033–50.

Drenth J, Jansonius JN, Koekoek R, Swen HM, Wolthers BG. Structure of papain. Nature 1968;218:929–32.

Ezcurra I, Ellerstrom M, Wycliffe P, Stalberg K, Rask L. Interaction between composite elements in the napA promoter: both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. Plant Mol Biol 1999;40:699–709.

Fischer J, Becker C, Hillmer S, Horstmann C, Neubohn B, Schlereth A, et al. The families of papain- and legumain-like cysteine proteinases from embryonic axes and cotyledons of *Vicia* seeds: developmental patterns, intracellular localization and functions in globulin proteolysis. Plant Mol Biol 2000;43:83–101.

Funk V, Kositsup B, Zhao C, Beers EP. The *Arabidopsis* xylem peptidase XCP1 is a tracheary element vacuolar protein that may be a papain ortholog. Plant Physiol 2002;128:84–94.

Gan S, Amasino RM. Inhibition of leaf senescence by autoregulated production of cytokinin. Science 1995;270:1986–8.

Garcia-Maroto F, Garrido-Cardenas JA, Michaelson LV, Napier JA, Alonso DL. Cloning and molecular characterisation of a Δ^8-sphingolipid-desaturase from *Nicotiana tabacum* closely related to Δ^6-acyl-desaturases. Plant Mol Biol 2007;64:241–50.

Gietl C, Schmid M. Ricinosomes: an organelle for developmentally regulated programmed cell death in senescing plant tissues. Naturwissenschaften 2001;88:49–58.

Grudkowska M, Zagdanska B. Multifunctional role of plant cysteine proteinases. Acta Biochim Pol 2004;51:609–24.

Gruis DF, Selinger DA, Curran JM, Jung R. Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases. Plant Cell 2002;14:2863–82.

Hensel LL, Grbic V, Baumgarten DA, Bleecker AB. Developmental and age-related processes that influence the longevity and senescence of photosynthetic tissues in arabidopsis. Plant Cell 1993;5:553–64.

Hori K, Watanabe Y. Context analysis of termination codons in mRNA that are recognized by plant NMD. Plant Cell Physiol 2007;48:1072–8.

Horsch RB, Fry JE, Hoffmann NL, Eichholtz D, Rogers SG, Fraley RT. A simple and general-method for transferring genes into plants. Science 1985;227:1229–31.

Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, et al. Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. Adv Bioinform 2008:420747.

Koizumi M, Yamaguchi-Shinozaki K, Tsuji H, Shinozaki K. Structure and expression of two genes that encode distinct drought-inducible cysteine proteinases in *Arabidopsis thaliana*. Gene 1993;129:175–82.

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. Protein measurement with the Folin phenol reagent. J Biol Chem 1951;193:265–75.

Muntz K, Belozersky MA, Dunaevsky YE, Schlereth A, Tiedemann J. Stored proteinases and the initiation of storage protein mobilization in seeds during germination and seedling growth. J Exp Bot 2001;52:1741–52.

Muntz K, Shutov AD. Legumains and their functions in plants. Trends Plant Sci 2002;7:340–4.

Nong VH, Becker C, Muntz K. cDNA cloning for a putative cysteine proteinase from developing seeds of soybean. Biochim Biophys Acta 1995;1261:435–8.

Ochman H, Ajioka JW, Garza D, Hartl DL. Inverse polymerase chain reaction. Biotechnology 1990;8:759–60.

Rzhetsky A, Nei M. Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. J Mol Evol 1992;35:367–75.

Schaller A. A cut above the rest: the regulatory function of plant proteases. Planta 2004;220:183–97.

Shindo T, Van Der Hoorn RAL. Papain-like cysteine proteases: key players at molecular battlefields employed by both plants and their invaders. Mol Plant Pathol 2008;9:119–25.

Smart CM. Gene-expression during leaf senescence. New Phytol 1994;126:419–48.

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24:1596–9.

Taylor MAJ, Baker KC, Briggs GS, Connerton IF, Cummings NJ, Pratt KA, et al. Expression of the pro-regions of papain and papaya proteinase IV in *Escherichia coli* and their inhibition of mature cysteine proteinases. Protein Eng 1995;8:59–62.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25:4876–82.

Vierstra RD. The ubiquitin-26S proteasome system at the nexus of plant biology. Nat Rev Mol Cell Biol 2009;10:385–97.

Wiederanders B. Structure–function relationships in class CA1 cysteine peptidase propeptides. Acta Biochim Pol 2003;50:691–713.

Wu J, Kang JH, Hettenhausen C, Baldwin IT. Nonsense-mediated mRNA decay (NMD) silences the accumulation of aberrant trypsin proteinase inhibitor mRNA in *Nicotiana attenuata*. Plant J 2007;51:693–706.