Korean J. Medicinal Crop Sci. 13(5) : 293 - 297 (2005)

Isolation and Characterization of Thioredoxin cDNA from *Codonopsis lanceolata* (S. et Z.) Trautv

Jun Gyo In*, Bum Soo Lee*, Yeong Deok Rho, Chang Yeon Yu***, and Deok Chun Yang**†**

**Biopia Co., Ltd., 383-4 Singal-ri, Kihung-eup Yongin, Kyunggi-do 449-598, South Korea.*
***Dept. of Oriental Medicinal Material & Processing, College of Life Science, Kyung Hee University, 1 Seocheon, Kihung-eup Yongin, Kyunggi-do 449-701, South Korea.*
****College of Agriculture and Life Science, Kangwon National University, Chunchon 200-701, South Korea.*

ABSTRACT : A thoredoxin (*CTRX*) gene was cloned and characterized from a full length cDNA library prepared from taproot of three-year old *Codonopsis lanceolata*. A *CTRX* was 666 nucleotides long and has an open reading frame of 372 bp with 124 amino acid residues (pI = 4.92). The deduced amino acid sequence of the *CTRX* matched to the previously reported plant thioredoxin *h* genes. The deduced amino acid sequence of *CTRX* exhibited the similarity of 33-67% among previously registered thioredoxin genes. The expression of *CTRX* in leaves of *Codonopsis lanceolata* was increased by wounding and 1 mM H_2O_2, but decreased by 0.1 mM cadmium.

Key words : Abiotic stress, cadmium, *Codonopsis lanceolata*, RT-PCR, thioredoxin

INTRODUCTION

Thioredoxin (Trx) is a ubiquitous small protein that is a subtrate for the NADPH-dependent seleno-flavoenzyme thiredoxin reductase. They constitute a family of small and ubiquitous proteins with two close and active Cys residues in a conserved motif: WCG/PPC. In their dithiol form, they are powerful disulfide reductases (Holmgren, 1985) that play a posttranslational regulatory role on protein targets involved in an ever-increasing number of cellular processes.

Plant have four different types of Trx isoforms, the well-known f and m types in the chloroplast (Buchanan, 1991), the recently described o type in the mitochondria (Laloi *et al.*, 2001), and the cytosolic *h* type possibly associated with plasma membrane (Johnson *et al.*, 1987a; Florencio *et al.*, 1988; Rivera-Madrid *et al.*, 1995).

Thioredoxin *h* is generally assumed to be cytosolic, which was supported by the absence of a transit peptide in the genes cloned for the isoforms from tobacco (Marty & Meyer, 1991), *Arabidopsis* (Rivera-Madrid *et al.*, 1993), *Triticum aestivum* (Gautier *et al.*, 1998), poplar (Balmer and Bunchana, 2002), germinating wheat seeds (Serrato *et al.*, 2001) and barley seed proteome (Kenji *et al.*, 2003). Moreover, the existence of several forms of thioredoxin *h* detected in spinach leaves (Florencio *et al.*, 1988), wheat flour (Johnson *et al.*, 1987b), and rice phloem sap (Ishiwatari *et al.*, 1995), supports the view that most higher plants possess multiple and divergent thioredoxin genes (Rivera-Madrid *et al.*, 1995). Whether these isoforms have different specificities or functions in the plant is not known.

In the present study, to obtain clues for possible diverse roles and functions in *C. lanceolata* taproots, a thioredoxin *h* cDNA was isolated and molecular characterization and expression analysis were conducted.

MATERIALS AND METHODS

RNA purification and cDNA library construction

Total RNA was isolated from 3-year taproot of *C. lanceolata* using aqueous phenol extraction procedure as described by Morris *et al.* (1990). The taproot was frozen and ground in liquid nitrogen prior to extraction of RNA. Poly $(A)^+$ RNA was isolated by oligo (dT) cellulose column using the Poly(A) Quick mRNA isolation kit (Stratagene, USA). A commercial cDNA synthesis kit was used to construct library according to the manufacture's instruction manual (Clontech, USA). To produce single-stranded cDNA appropriate for directional cloning 5 μg of poly $(A)^+$ RNA was primed with an oligo (dT) primer. Double-stranded cDNA was produced using RNase H and *E. coli* DNA polymerase. After ds cDNA synthesis by primer extension, cDNA was digested with *Sfi* I and then size-fractionated through a Chroma spin-400 column. Size-selected cDNA was ligated into λTriplEx2 vector and was packaged *in*

† Corresponding author: (Phone) +82-31-201-2688 (E-mail) dcyang@khu.ac.kr
Received September 11, 2005 / Accepted October 20, 2005

vitro using Gigapack III Gold Packaging Extract kits (Stratagene, USA). Fractions containing cDNA greater than 500 bp were recovered and this library was amplified once to yield a final titre of 2×10^9 pfu $m\ell^{-1}$.

Nucleotide sequencing and sequence analysis

pTriplEx2 phagemids were converted from λTriplEx2 library in *E. coli* strain BM25.8. Phagemids containing inserts were selected by blue and white color screening on IPTG/X-GAL/ampicillin plates. Single-run partial sequencing of such randomly selected cDNA clone was performed. The 5' ends of the cDNA inserts were sequenced using sequencing primer by an automatic DNA sequencer (ABI prism 3700 DNA sequencer, Perkin-Elmer, USA) according to the thermal cycling protocol of the BigDye Terminator Cycle Sequencing kit. Sequences were edited to remove the vector sequence, poly A tails and ambiguous regions. Bacterial genomic DNA sequences were identified by BLASTN comparisons against the GenBank non-redundant databases. Sequences shorter than 100 bases were discarded. The individual ESTs were searched against the GenBank nr database using a BLASTX algorithm. A pTriplEx phagemid for thioredoxin cDNA was excised from the λTriplEx2 and used as templates for sequence analysis. The cDNA insert was sequenced using the 5' and 3' sequencing primer by an automatic DNA sequencer (ABI prism 3700, USA). Nucleotide and amino acid sequence analyses were performed using DNASIS program (Hitachi, Japan).

Plant materials and abiotic stress treatments

Three-year old *C. lanceolata* grown at field were used. Leaves attached in the upper region cut with a scalpel. Cut leaves were soaked in 1 mM H_2O_2 for the oxidative stress or in 100 uM cadmium ($Cd\cdot5/2H_2O$, Sigma) during 30 min, 1 hr, 2 hr, and 4 hr, respectively. After the stress treatments, the leaves were immediately frozen in liquid nitrogen and stored at −80℃.

Total RNA preparation

Total RNA was isolated from the stress treated *C. lanceolata* leaves using the method of guanidine isothiocyanate (TRIzol, Gibco BRL). The tissue was frozen with liquid nitrogen and ground to a fine powder with a mortar and pestle. Extraction reagent was added 1 mℓ and ground more than 3 min. The mixture was centrifuged for 5 min at 4℃. The supernatant was removed to a new tube, added 0.2 mℓ chloroform, and vigorously mixed. Following centrifugation, total RNA was precipitated with isopropanol. The pellet was washed once with 75% ethanol, dried in vacuum for 3 min and dissolved in DEPC treated DDW. The total RNA solution was stored at −80℃ for the RT-PCR analysis.

Quantitative RT-PCR analysis

To analyze the gene expression of *CTRX* gene against the oxidative and heavy metal stresses, we employed the quantitative RT-PCR. A pair of gene-specific primers of the *CTRX* gene was designed and used for RT-PCR analysis. Specific primers included the following: (*CTRX*-forward) 5'-ATA TCG AAA GTA GTG AAG GTG GAC TC-3'; (CTRX-reverse) 5'-GGC CAC ACA AAT GGG ATT GGA CTG T-3'. As a control, we used a pair of the specific primers to *C. lanceolata* actin gene (CAct), 5'-CGA GAA GAG CTA CGA GCT ACC CGA TGG-3' (forward) and 5'-CTC GGT GCT AGG GCA GTG ATC TCT TTG CT-3' (reverse). Ten microgram of total RNA was used for the RT-PCR analysis, according to the method of Takakura *et al.* (2000). The PCR cycles for quantitative RT-PCR numbered 30 for the *CTRX* and the *PAct* genes. RT-PCR products were run on 1.5% (W/V) agarose gel in 0.5×TAE buffer and then photograped for the expression analysis.

RESULTS AND DISCUSSION

To isolate useful genes against various abiotic stresses from *C. lanceolata*, we analyzed 1,000 EST clones randomly

```
AAACTAAATTAAAAGTAAAAGAAGAATCAGGGCCAGGATAGAGCAAGAACAGCAACAAGA    60
AGGAGATGGAACATCATCAAATTAGCATATCGAAAGTAGTGAAGGTGGACTCAAAGCAAA   120
    M  E  H  H  Q  I  S  I  S  K  V  V  K  V  D  S  K  Q  T    19

CTTGGGATTCTCTTGTAATTCAAGCCATTAATCAAAGCTCCCCTATTGTGGCACATTTTA   180
  W  D  S  L  V  I  Q  A  I  N  Q  S  S  P  I  V  A  H  F  T    39

CAGCATCATGGTGCATTCCATCGGTTGTGATGAACCCCTTCTTCGAAGAGTTGGCCTCGA   240
  A  S  W  C  I  P  S  V  V  M  N  P  F  F  E  E  L  A  S  T    59

CTTACAAAGACACAATGTTTCTCTTGGTGGATGTGGATGATGTTCAGGAGGTCGCGAGCA   300
  Y  K  D  T  M  F  L  L  V  D  V  D  D  V  Q  E  V  A  S  K    79

AGTATGAAGTCAAGGCAATGCCAACATTTGTGCTGATGAAGGAAGGAGCTCAGGTTGACA   360
  Y  E  V  K  A  M  P  T  F  V  L  M  K  E  G  A  Q  V  D  K    99

AGCTAATTGGTGCAAATCCAGATGAGATAAGGAAAAGGATTGAGTCTCTTATACAGTCCA   420
  L  I  G  A  N  P  D  E  I  R  K  R  I  E  S  L  I  Q  S  N   119

ATCCCATTTGTGTGGCCTAACTTAGCCTAGCCATAGGACTCGGAGCGTGTGTTTTTTTTT   480
  P  I  C  V  A  *                                             124

TCCCCACATTTCCAAATTCCGACGATATGATTGTGGCTGTATATTAATATTGCTCCGATT   540
ATGTATGATATATGTATGATAGCAGTGGATCTTCATAGTTGGGGGTTTTTGTACAGAAAG   600
ATGTTGGGGACCGCTAGACGGTTAGCTTTTTTTGCAGATTGAGCTCGAGTNCCAGTTTTC   660
TATCTT                                                         666
```

Fig. 1. Nucleotide and deduced amino acid sequence of *CTRX*, thioredoxin cDNA, from *Codonopsis lanceolata*. The positions of nucleotides are shown on the left and the positions of amino acids under the below. Asterisk shows the termination codon. The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequencing Database under the accession number AB223035.

```
Codonopsis      -MEHHQISISKVVKVDSKQTWDSLVIQAINQSSPIVAHFTASWCIPSVVMNPFFEELAST  59
Pisum           --MAGSSEEGQVISCHSVDAWNDILHRGNESKKLIVVDFTASWCGPCRFIAPFLGELAKK  58
Citrus          --MA---EEGQVISCHTVESWNEQLQKGIAAKKLIVVDFTASWCPPCKLMSPILSELAKK  55
Fagopyrum       --MA---EEAQVIACHTVQEWNEKFQKAKDSGKLIVIDFTASWCGPCRVITPYVSELAKK  55
Triticum        --MA---AEEGAVIACHTKQEFDTHMANGKETGKLVIIDFTASWCGPCRVIAPVFAEYAKK  56
Oryza           --MA---AEEGVVIACHNKDEFDAQMTKAKEAGKVVIIDFTASWCGPCRFIAPVFAEYAKK  56
Cucurbita       --MA----VEGQVIPLRDAEEFDAIIDKEKESGRLIVIDFTASWCPPCRFIAPVFAELAKA  55
Helicosporidium --MG------GAVIQVTNKTEFDSHLSTAASKGKLVVVDFTATWCGPCKMIAPFFAKLSGE  53
Hevea           METQEQQAKSRVVKVDSVESWDFYVTQANNGGCPIVVHFTASWCIPSVAMNPFFEELASA  60
Arabidopsis     ---------MARVVKIDSAESWDFYVSQAKNQNCPIVAHFTALWCIPSVFMNSFFEELAFN  52
                         . *:      :: .      :: .*** ** *. : . . : :

Codonopsis      YKDTMFLLVDVDDVQEVASKYEVKAMPTFVLMKEGAQVDKLIGANPDEIRKRIESLIQSN 119
Pisum           FTNVIFLKVDVDELKSVAQDWAVEAMPTFVFVKEGTILGKVVGAKKEELQQTIERHVAS- 117
Citrus          LPAVIFLKVDVDELQSVAEEWAVEAMPTFVLTKDGKVLERIVGAKKDELQLAVEKHATTV 115
Fagopyrum       FPHVAFFKVDVDDLKDVAEEYKVEAMPSFVILKEGQEVERIVGARKDELLHKIAVHAPI- 114
Triticum        FPGAIFLKVDVDELKDVAEAYNVEAMPTFLFIKDGAKVDTVVGGRKDDIHTKIVALMG--- 114
Oryza           FPGAVFLKVDVDELKEVAEKYNVEAMPTFLFIKDGAEADKVVGARKDDLQNTIVKHVGA- 115
Cucurbita       HVHVTFLKVDVDNVKEIAKRFEVTAMPTFVFLKGGNEVHRIVGADKVELGVKVLELSAAP 115
Helicosporidium YPDVVFLKVDVDEVEAVAAEHGITAMPTFLFFKDGKQVDSLTGANQERLRAMLTQHAQS- 112
Hevea           YPDVLFLAVDVDEVKEVASKLEVKAMPTFVLMKDGAQIDRLVGANPEEIRKRIGGFAQSI 120
Arabidopsis     YKDALFLIVDVDEVKEVASQLEVKAMPTFLFLKDGNAMDKLVGANPDEIKKRVDGFVQSS 112
                   . *: ****::: :*    : ***:*:: * *     : *.   :   :

Codonopsis      ----PICVA- 124
Pisum           -------SNA- 120
Citrus          ENATTANA- 123
Fagopyrum       -------TA- 116
Triticum        -----SASA- 118
Oryza           -TAASASA- 122
Cucurbita       -----ATCAS 120
Helicosporidium ---------- 
Hevea           ----RVAVA- 125
Arabidopsis     ----RVVHIA 118
```

Fig. 2. Multiple alignment of the deduced amino acid sequence of *CTRX* with those of thioredoxin genes from other plants; *C. lanceolata* (AB223035), *H. brasiliensis* (AAD33596), *A. thaliana* (AAM63200), *P. sativum* (CAC 36986), *T. aestivum* (AAL67139), *C. maxima* (BAC21264), *H. sp.* ex *Simulium jonesii* (AAU93947), F. *esculentum* (Q96419), *O. sativa* (XP_476912), and C. x *paradisi* (AAP33009). Sequence data was obtained from GeneBank listed and aligned using DDBJ ClustalW (Thompson *et al.*, 1994 and 1997) and GeneDoc (Nicholas *et al.*, 1997).

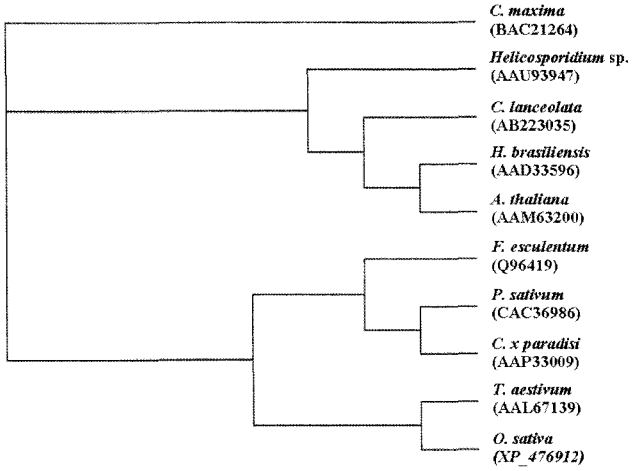

Fig. 3. Phylogeny of the thioredoxin protein family from *C. lanceolata* and other plants. Phylogenetic analysis is based on the deduced amino acid sequences of thioredoxin genes from various plant species. The branch lengths are proportional to divergence, with the scale of 0.1 representing 10% change.

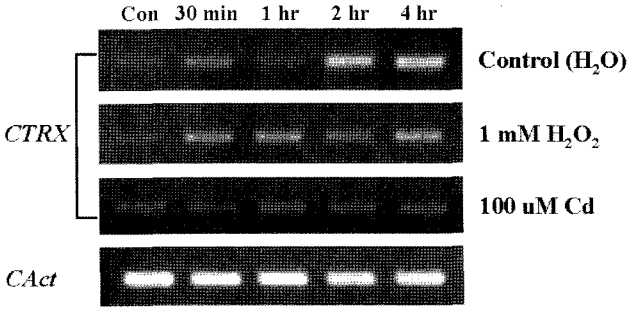

Fig. 4. Expression pattern of *CTRX* under the stress of cadmium and hydrogen peroxide. Total RNA from stress-treated samples served as templates for quantitative RT-PCR with gene-specific primers. Actin gene isolated from *C. laceolata* was used for RT-PCR control.

selected. A thioredoxin *h* homolog of analyzed EST clones was isolated and named as *CTRX*. *CTRX* cDNA was 666 nucleotides long and possess an open reading frame of 372 bp with 65 bp 5'-untranslated region (5'UTR) and 226 bp 3'UTR (Fig. 1). BLASTX database searches with the *CTRX* sequence gave thioredoxins from various organism as the top 50 best matches. The *CTRX* cDNA encodes a protein of 12 amino acids (MW 13.9 kDa) with pI 4.92 (Fig. 2).

The deduced amino acid sequence of *CTRX* cDNA was compared to other related thioredoxin gene products registerd previously (Fig. 2). A *CTRX* protein sequence showed 33-67% similarities with one another. They have no transit peptide, suggesting that they are located in the cytosol. Fig. 2 shows a multiple alignment of the *C. lanceolata* thioredoxin *h* protein and other plant thioredoxin *h* homologous molecules available in the GenBank. The conserved region contains the redox-active cysteines in the consensus sequence WCGPC. This motif is almost completely conserved in all thioredoxins except some prokaryotic sequences.

A comparison of the *CTRX* sequences with the nine Trx *h* sequences in databases revealed that it was more similar to those of Arabidopsis and Hevea than to any other species (data not shown). A phylogenetic tree was constructed with the *CTRX* and other plant thioredoxins h. The *CTRX* was found in the same subgroup with *H. brasiliensis* and *A. thaliana* (Fig. 3).

We performed quantitative RT-PCR for the analysis of *CTRX* gene expression against abiotic stresses, such as hydrogen peroxide (H_2O_2) and cadmium. For the confirmation of the *CTRX* gene, we have cloned and sequenced the PCR product amplified with *CTRX* specific primers. Expression of the actin gene isolated from *C. lanceolata* served as an internal control for quantitative RT-PCR analysis.

C. lanceolata leaves attached petiole were soaked in 1 mM hydrogen peroxide. Total RNA was extracted from each samples and performed the RT-PCR analysis (Fig. 4). We used distilled water as a control. The *CTRX* gene was slightly induced by 1 mM hydorgen peroxide than control. When *Codonopsis*

leaves cut using scalpel were soaked in water, the expression of *CTRX* gene was increased from 2 hr to 4 hr after treatment (Fig. 4). The Arabidopsis cytosolic system consists of eight different thioredoxins *h* and two homodimeric NADPH-dependent thioredoxin reductases (Laloi *et al.*, 2001; Meyer *et al.*, 2002). Recently, Laloi *et al.* (2004) reported that expression of the Arabidopsis cytosolic thioredoxin *h*5 gene was closely related to wounding, abscission, senescence, and pathogen attack, as well as several different oxidative stress conditions.

In order to investigate the response of *CTRX* gene against cadmium, leaves of *C. lanceolata* were soaked in 100 uM Cd. But, the transcription of *CTRX* gene was not activated by the cadmium treatment (Fig. 4). TRX proteins can coordinate heavy metals by thiol bonds and confer heavy metal tolerance on various organisms. Consequently, the active site dithiol of reduced TRXs is a possible target of heavy metal fixation on the protein. Also, the stress induced by heavy metals might be related to an oxidative stress. In *Chlamydomonas reinhardtii*, Lemaire *et al.* (1999) have reported that the expression of the chloroplastic thioredoxin *m* and one cytosolic thioredoxin *h* is up-regulated by the heavy metals Cd and Hg. Because treatments with oxidants did not trigger any accumulation of TRX messengers, Lemaire *et al.* (1999) concluded that the induction of TRX expression by heavy metals was a direct response to the cations as suggested by the presence of cis-acting elements related to cadmium induction in the TRX *h* promoter. In contrast, Laloi *et al.* (2004) have reported that the activation of the Arabidopsis *AtTRXh5* gene by oxidants such as menadione as well as the absence of cis-acting elements related to heavy metal induction in the *AtTRXh5* promoter are in favor of an activation of expression mediated by ROS. The expressionn of *CTRX* gene in leaf of *C. lanceolata* was increase by wounding and hydorgen peroxide, but not cadmium. This suggests that among the thioredoxin *h* family, the *CTRX* might be specifically involved in resistance against wounding stress.

We described here the isolation and characterization of thioredoxin *h* gene in *C. lanceolata*, *CTRX*. The functions of TRX *h* in plants are still largely unknown. Therefore, further studies are need to understand the physiological function of TRX regulation in *C. lanceolata*.

ACKNOWLEDGEMENTS

This work was supported in part by the BIOGREEN 21 Program.

LITERATURE CITED

Balmer Y, Buchanan BB (2002) Yet another plant thioredoxin. Trends Plant Sci. 7:191-193.

Buchanan BB (1991) Regulation of CO_2 assimilation in oxygenic photosynthesis:the ferredoxin/thioredoxin system. Perspective on its discovery, present status, and future development. Arch. Biochem. Biophys. 288:1-9.

Florencio FJ, Yee BC, Johnson TC, Buchanan BB (1988) An NADP thioredoxin system in leaves: purification and characterization of NADP thioredoxin reductase and thioredoxin *h* from spinach. Arch. Biochem. Biophys. 26:496-507.

Gautier MF, Lullien-Pellerin V, De Lamotte-Guery F, Guirao A, Joudrier P (1998) Characterization of wheat thioredoxin *h* cDNA and production of an active *Triticum aestivum* protein in *Escherichia coli*. Eur. J. Biochem. 252:314-324.

Holmgren A (1985) Thioredoxin. Annu. Rev. Biochem. 54:237-271.

Ishiwatari Y, Honda C, Kawashima I, Nakamura S, Hirano H, Mori S, Fujiwara T, Hayashi H, Chino M (1995) Thioredoxin *h* is one of the major proteins in rice phloem sap. Planta 195:456-463.

Johnson TC, Cao RQ, Kung JE, Buchanan BB (1987a) Thioredoxin and NADP-thioredoxin reductase from cultured carrot cells. Planta 171:321-331.

Johnson TC, Wada K, Buchanan BB, Holmgren A (1987b) Reduction of purothionin by the wheat seed thioredoxin system. Plant Physiol. 85:446-451.

Kenji M, Finnie C, Østergarrrd O, Svensson B (2003) Identification, cloning and characterization of two thioredoxin *h* isoforms, HvTrxh1 and HvTrxh2, from the barley seed proteome. Eur. J. Biochem. 270:2633-2643.

Laloi C, Rayapuram N, Chartier Y, Grienenberger JM, Bonnard G, Meyer Y (2001) Identification and characterization of a mitochondrial thioredoxin system in plants. Proc. Natl. Acad. Sci. USA 98:14144-14149.

Lemaire S, Keryer E, Stein M, Schepens I, Issakidis-Bourguet E, Gerard-Hirne C, Miginiac-Maslow M, Jacquot J (1999) Heavy-mtal regulation of thioredoxin gene expression in *Chlamydomonas reinhardtii*. Plant Physiol. 120:773-778.

Laloi C, Rayapuram N, Chartier Y, Grienenberger JM, Bonnard G, Meyer Y (2001) Identification and characterization of a mitochondrial thioredoxin system in plants. Proc. Natl Acad. Sci. USA 98:14144-14149.

Laloi C, Mestres-Ortega D, Marco Y, Meyer Y, Reichheld JP (2004) The arabidopsis cytosolic thioredoxin *h*5 gene induction by oxidative stress and its W-box-mediated response to pathogen elicitor. Plant Physiol. 134:1006-1016.

Marty I, Meyer Y (1991) Nucleotide sequence of a cDNA encoding a tobacco thioredoxin. Plant Mol. Biol. 17:143-147.

Meyer Y, Vignols F, Reichheld JP (2002) Classification of plant thioredoxins by sequence similarity and intron position. Methods Enzymol. 347:394-402.

Morris PC, Kumar A, Bowles DJ, Cuming AC (1990) Osmotic stress and abscisic acid regulate the expression of the *Em* gene of wheat. Eur. J. Biochem. 190:625-630.

Nicholas KB, Nicholas HB (1997) GeneDoc: analysis and visualization of genetic variation. http://www.cris.com/~Ketchup/genedoc.html.

Rivera-Madrid R, Marinho P, Brugidou C, Chartier Y,

Meyer Y (1993) Nucleotide sequence of a cDNA clone encoding an *Arabidopsis thaliana* thioredoxin *h.* Plant Physiol. 102:327-328.

Rivera-Madrid R, Mestres D, Marinho P, Jacquot JP, Decottignies P, Miginiac-Maslow M, Meyer Y (1995) Evidence for five divergent thioredoxin *h* sequences in *Arabidopsis thaliana.* Proc. Natl. Acad. Sci. USA 92:5620-5624.

Serrato AJ, Crespo JL, Florencio FJ, Francisco JC (2001) Characterization of two thioredoxins *h* with predominant localization in the nucleus of aleurone and scutellum cells of germinating wheat seeds. Plant Mol. Biol. 46:361-371.

Takakura Y, Ito T, Saito H, Inoue T, Komari T, Kuwata S (2000) Flower-predominant expression of a gene encoding a novel class I chitinase in rice (*Oryza sativa* L.). Plant Mol. Biol. 42:883-897.

Thompson JD, Higgins DG, Gilbson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-secific gap penalties and weight matrixchoice. Nucleic Acids Res. 22:4673-4680.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882.