Cloning and expression analysis of cysteine protease gene (*MwCP*) in *Agropyron mongolicum* Keng

T.G.B.Y. Ao[1,3], M.L. Lang[2], Y.Q. Li[3], Y. Zhao[1], L.C. Wang[1] and X.J. Yang[1,2,4]

[1]College of Agronomy, Agricultural University of Hebei, Baoding, China
[2]College of Life Sciences, Agricultural University of Hebei, Baoding, China
[3]College of Animal Science and Technology, Agricultural University of Hebei, Baoding, China
[4]Key Laboratory of Crop Germplasm in Hebei Province, Agricultural University of Hebei, Baoding, China

Corresponding author: X.J. Yang
E-mail: yangxueju2015@163.com

ABSTRACT. In this study, a cysteine protease gene (*MwCP*) from *Agropyron mongolicum* Keng was isolated using RACE. Sequence analysis indicated that *MwCP* was 1473 bp, and it contained a 1134-bp open reading frame, which encoded 377 amino acids with a 24-amino acid N-terminal signal peptide. The results indicated that the MwCP protein was a new member of the papain C1A family, and it was predicted to be an extracellular, secretory stable hydrophilic protein. The secondary structure of MwCP was mainly composed of α-helices and random coils, and the space structure primarily contained α-helices, β-sheets, and β-turns. Homology analyses showed the 98% homology between MwCP amino acids and a cysteine protease found in *Triticum aestivum* (GenBank accession No. AAW21813.1). Analysis of mRNA using semi-quantitative RT-PCR indicated that during a 48-h drought stress period, *MwCP* was expressed during the 4th hour, and the expression level peaked during the 6th hour before declining to

the original level. The results revealed that *MwCP* was involved in drought-resistant physiological processes of *A. mongolicum*. Moreover, the *MwCP* expression levels were highest in leaves, intermediate in roots, and lowest in stems.

Key words: *Agropyron mongolicum* Keng; Cysteine protease gene; Gene cloning; Expression analysis

INTRODUCTION

Agropyron mongolicum Keng (Gramineae) is a precious diploid species and forage type that is found in desert grasslands and typical grasslands of China (Zhao, 2009). *A. mongolicum* is a valuable forage resource that is resistance to drought, sand storms, and barrenness. It also has several tolerance genes that contribute to crop resistance or tolerance, so it can be used to prevent wind, fix drifting sands, and conserve water and soil (Li and Dong, 1993; Yamaguchi-Shinozaki and Shinozaki, 1993; Griffiths et al., 1997; Zhang et al., 2004). *A. mongolicum* was listed as a national second-class rare and endangered plant, and it is a wild relative of crops that need protection (Che and Yang, 2010). Reports regarding the genetic engineering of drought-resistant *A. mongolicum* have gradually increased in recent years (Cui, 2000; Zhao et al., 2008).

Cysteine protease (CP) is a type of protease that is associated with plant stress physiology. As an important protease family, CPs are widely involved in plant physical processes such as anabolic and catabolic processes, programmed cell death, tissue aging, and seed development and constitution when cells are subjected to environmental stresses (Li, 2008; Yuan et al., 2008; Wang et al., 2012). For instance, barley (*Hordeum vulgar*) has 42 types of proteases that are involved in germination, and 27 of these are CPs (Grudkowska and Zagdańska, 2004). The importance of CPs in plants is obvious, and different CP forms are activated to degrade inactive proteins when plants are subjected to the stresses of drought, disease, and pests (Yan et al., 2005; Zang et al., 2010). Studies have confirmed that during programmed cell death (PCD), CP and endonuclease are two catalytic protein types that are commonly involved in the process. Furthermore, the CP family directly contributes to the protoplast disintegration of cells during PCD (Cui, 2000; Mohan et al., 2006; Chen et al., 2006; Aberlenc-Bertossi et al., 2008; Zhu et al., 2009).

Environmental stresses such as drought can greatly damage crop production (Stolf-Moreira et al., 2010). Several researchers have conducted systematic research on the morphological, physiological, and molecular characteristics of drought-resistant plants (Nassar et al., 2010; Reis et al., 2014). In this study, the RACE method was used to isolate the *A. mongolicum* CP genes that responded to drought stress. The *MwCP* gene might enrich forage crop gene pools that are related to drought-resistance, and it could also provide basic data for molecular studies of drought resistance in forage and gramineous crops.

MATERIAL AND METHODS

Materials and reagents

A. mongolicum seeds were provided by the Forage Experiment Station of Inner Mongolia University. Competent *Escherichia coli* cells Trans-T1, cloning vector PEASY-T1, and TransFast

Taq DNA Polymerase were purchased from Transgen Biotech Company. A total RNA extraction kit and a TIANScript RT kit for first-strand synthesis were obtained from TIANGEN Biotech Co. Ltd. A SMART RACE cDNA Amplification Kit (Clontech) was purchased from TaKaRa Company. Other reagents, including a DNA gel extraction kit, IPTG, and X-gal were purchased from Sangon Biotech Co. Ltd. (Shanghai), and this company also conducted primer synthesis and sequencing.

Material preparation

A. mongolicum seeds were washed in water and soaked in water to stimulate germination. The seeds were then sowed in trays with vermiculite before being placed in the illumination box. Seeds were cultured under the following conditions: 30-day cultivation of 16 h in light and 8 h in darkness per day in 28°C. Seedlings were then harvested, and were subsequently treated with 20% PEG-6000 to simulate 6 h of drought treatment in order to induce the expression of related genes. The samples were then quickly frozen with liquid nitrogen, and the seedlings were stored in at -70°C.

Primer design

Using both DNAMAN and Primer Premier 5.0 programs, the conserved sequences encoded by plant CP genes (in GenBank) were used to design the required primers (Table 1). The working concentration of each primer was 10 mol/μL.

Table 1. Primers used to clone *Agropyron mongolicum* Keng *MwCP* cDNA.

Primer name	Primer sequence (5'-3')	Sequence length (bp)
T7dT18	ACGACTCACTATAGGGC(dT)$_{23}$	40
CP-F2	TGGATCATCAAGAACTCATG	20
OligodT12	ACGACTCACTATAGGGCTTTTTTTTTTTT	29
UPM	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	45
GSP-1	GGCTTCGCTCGTACCGCAGCGGCAAG	26
NUP	AAGCAGTGGTATCAACGCAGAGT	23
GSP-2	GACCAGAGCCTACTCCTCCT	20
GSP-3	TTCATCAACAGCGACAACGCTG	22
CP-FSac2	AATTCCCCAATGGATCATCGCCTCGTGGC	26
CP-RXho2	CCGCTCGAGCTACTCCTCCTTGGAGGCAT	29
A4-F	CCGTTCTGTCCTTGTATGCCA	21
A4-R	CCACATCTGCTGGAAAGTGC	20

Extraction and reverse transcription of RNA

Using ground materials, RNA extraction and RT-PCR analyses were performed according to the manufacturer instructions, with some modifications. Regarding the synthesis of 3'-RACE ready cDNA, the Oligo dT primer provided in the TIANScript RT kit was replaced with T7dT18. The manufacturer protocols of the SMART RACE cDNA Amplification kit (Clontech) were referred to during the synthesis of first-strand cDNA using 5'-RACE.

3'-RACE

The CP-F2 primer (Table 1), used for the amplification of the 3'-end of the plant CP gene,

was designed based on previously reported conserved CP sequences. RT-PCR was conducted to amplify the 3'-end of the sequence with the first-strand *A. mongolicum* cDNA serving as the template. CP-F2 was used as the upstream primer, and the adapter OligodT12 primer of the 3'-end was used as the downstream primer. The total 3'-RACE PCR volume was 25 μL, and it contained 1 μL 3'-RACE ready cDNA, 1 μL primer (CP-F2 and OligodT12), and 12.5 μL mix (ddH$_2$O was added to bring the total volume to 25 μL). The reaction conditions were as follows: pre-denaturing at 94°C for 5 min; 30 denaturing cycles at 94°C for 30 s, annealing at 48°C for 1 min, and extension at 72°C for 1 min; and a final extension at 72°C for 10 min. We detected RT-PCR products via 2% agarose gel electrophoresis. The target bands were collected, and were then cloned into the PEASY-T1 vector before being sent off-site for sequencing.

5'-RACE

Amplification of a portion of the 5'-end sequence

The GSP-1 and GSP-2 primers (Table 1), used for amplification of the 5'-end of the sequence, were designed based on the conserved region of the 3'-end sequence obtained from 3'-RACE analysis. Nested PCR was conducted with the UPM and NUP 5'-end adapter primers (Table 1) according to the SMART RACE cDNA Amplification kit instructions (Clontech). The total reaction volume of the first round of PCR was 25 μL, and it consisted of 1.2 μL 5'-RACE ready cDNA, 0.5 μL GSP-1, 2.5 μL UPM, and 12.5 μL mix (ddH$_2$O was added to bring the total volume to 25 μL). The reaction conditions were as follows: pre-denaturing at 94°C for 5 min; five denaturing cycles at 94°C for 30 s, and extension at 72°C for 3 min; five denaturing cycles at 94°C for 30 s, annealing at 70°C for 45 s, and extension at 72°C for 3 min; 25 cycles of denaturing at 94°C for 30 s, annealing at 68°C for 45 s, and extension at 72°C for 3 min; and a final extension at 72°C for 12 min. The total reaction volume for the second round of PCR was 25 μL, and it included 0.1 μL PCR products obtained from the first round, 0.5 μL GSP-2, 0.5 μL NUP, and 12.5 μL mix (ddH$_2$O was added to bring the total volume to 25 μL). The reaction conditions were as follows: pre-denaturing at 94°C for 4 min; 30 denaturing cycles at 94°C for 30 s, annealing at 58°C for 45 s, and extension at 72°C for 1.2 min; and a final extension at 72°C for 12 min.

Amplification of the total 5'-end sequence

A GSP-3 primer was designed to continue amplifying the 5'-end of the sequence obtained from the above analysis, and GSP-1, UPM, and NUP were used to conduct the second round of nested PCR. The conditions of the first round of PCR were the same as those previously described. The second round of PCR (25 μL total volume) consisted of 0.1 μL PCR products obtained from the first round, 0.5 μL GSP-3, 0.5 μL NUP, and 12.5 μL mix (ddH$_2$O was added to bring the total volume to 25 μL). The reaction conditions of the second round were the same as previously described.

Acquisition of MwCP and the open reading frame (ORF)

After obtaining sequencing results, we assembled the cDNA sequence and analyzed it using the DNAMAN and DNAStar programs. The Primer Premier 5.0 program was used to design

the CP-FSac2 and CP-RXho2 primers (Table 1), which were used to amplify the target ORF via PCR. The total reaction volume was 25 μL, and the RT-PCR contained 1 μL 3'-RACE ready cDNA, 0.5 μL CP-FSac2, 0.5 μL CP-RXho2, and 12.5 μL mix (ddH$_2$O was added to bring the total volume to 25 μL). The reaction conditions were as follows: pre-denaturing at 94°C for 5 min; eight cycles of denaturing at 94°C for 30 s, annealing at 68°C for 60 s, and extension at 72°C for 1.2 min; and 28 cycles of denaturing at 94°C for 30 s, annealing at 65°C for 60 s, and extension at 72°C for 1.2 min; and a final extension at 72°C for 10 min.

Bioinformatic analysis

Homology analysis of amino acid sequence

The ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was used to search for an ORF in the cDNA sequence, which was then translated into an amino acid sequence. The sequence homology comparison was conducted using BLAST (NCBI; http://blast.ncbi.nlm.nih.gov/Blast.cgi). The DNAMAN 8.0 software was used for multiple-sequence alignment of homologous amino acid sequences. The MEGA 6.0 software was used to construct the phylogenetic tree using the neighbor-joining algorithm, and 1000 bootstrap replicates were used to statistically assess nodal support.

Structure prediction

The secondary structure of the coded protein was predicted using the online software SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) and the online application server SMART (http://smart.embl-heidelberg.de/). Conserved domains were analyzed using the NCBI Conserved Domain Database, and the three-dimensional structure of the coded protein was predicted using the software from the SWISS-MODEL homology model server (http://www.swissmodel.expasy.org/).

Physical and chemical property analysis

Basic physical and chemical properties of the coded protein, such as molecular weight, theoretical isoelectric point (pI), etc., were analyzed using the ProtParam program (http://web.expasy.org/protparam/). Moreover, a Kyte and Doolittle hydropathic plot was generated using the ProtScale program (http://web.expasy.org/protscale/).

Functional analysis

Signal peptides were predicted using the SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/), TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/), and iPSORT (http://ipsort.hgc.jp/) programs. Possible subcellular localization was calculated using the PSORT (http://psort.hgc.jp/form.html) and Wolf PSORT (http://www.genscript.com/psort/wolf_psort.html) programs. The PROSITE database (http://prosite.expasy.org/prosite.html) was utilized to analyze functional sites, and TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) was used to predict transmembrane regions.

Expression analysis of *MwCP*

Expression analysis after varied stress duration

In order to investigate *MwCP* involvement in drought resistance, we cultured *A. mongolicum* for 30 days, harvested the seedlings (see the "Material preparation" section), and soaked the seedlings in 20% PEG-6000 solution. The seedlings were then exposed to drought stress for 0, 2, 4, 6, 8, 10, 12, 24, 36, and 48 h prior to whole seedling collection. The collected materials were quickly frozen with liquid nitrogen and stored at -70°C. We extracted the total RNA of the samples, and synthesized cDNA via reverse transcription using equal RNA quantities (see the "Extraction and reverse transcription of RNA" section). We conducted RT-PCR with equal cDNA quantities, and CP-FSac2 and CP-RXho2 served as primers (see the "Acquisition of *MwCP* and the ORF" section). The *Actin* gene was simultaneously amplified using equal cDNA quantities, and A4-F and A4-R (Table 1) served as inner reference primers. The total volume of the Actin PCR was 25 μL, and it contained 1 μL cDNA, 0.5 μL A4-F, 0.5 μL A4-R, and 12.5 μL mix (ddH$_2$O was added to bring the total volume to 25 μL). The reaction conditions were as follows: pre-denaturing at 94°C for 3 min; 30 cycles of denaturing at 94°C for 30 s, annealing at 57°C for 45 s, and extension at 72°C for 1 min; and a final extension at 72°C for 10 min. Following RT-PCR, we detected PCR products using 2% agarose gel electrophoresis. We then performed quantitative analyses to determine the level of gene expression as compared to the expression level (as percentage values) of the *Actin* gene using the Gel-Pro software, which utilizes the gray scale of the visualized bands.

Expression analysis in different tissues

We cultured *A. mongolicum* for 30 days, harvested the seedlings (see the "Material preparation" section), and treated them with 20% PEG-6000 solution. We then simulated drought for 6 h prior to collecting the whole seedlings. The collected materials were quickly frozen with liquid nitrogen and stored at -70°C. We extracted total RNA from roots, stems, and leaves, and we then conducted RT-PCR using the same RNA quantities for conversion to cDNA (see the "Extraction and reverse transcription of RNA" section). We conducted RT-PCR with equal cDNA quantities, and CP-FSac2 and CP-RXho2 served as primers (see the "Acquisition of *MwCP* and the ORF" section). Moreover, the *Actin* gene was amplified using the previously described methods. Following RT-PCR, we detected PCR products using 2% agarose gel electrophoresis. We then performed quantitative analyses to determine the level of gene expression as compared to the expression level (as percentage values) of the *Actin* gene using the Gel-Pro software, which utilizes the gray scale of the visualized bands.

RESULTS

3'-RACE *MwCP* results

The results of the multiple-sequence alignment of the 3'-RACE amplification products (Figure 1a) indicated that the target fragment, including a poly-A tail, had relatively high homology (>85%) with the 3'-end sequence of several types of CP genes from plants. This result indicated that the target fragment was the 3'-end of the *MwCP* sequence.

5'-RACE *MwCP* results

The results of the multiple sequence alignment of the 5'-RACE amplification products (Figure 1b) indicated that the target fragment, which was a portion of the 5'-end of the *MwCP* sequence, had relatively high homology (>95%) with the 5'-end sequence of several types of CP genes from plants. Therefore, we designed the GSP-3 primers used for a second nested PCR, which were based on the products previously obtained from the first round of PCR, and an 827-bp fragment was observed (Figure 1c). The multiple-sequence alignment indicated that the fragment was the complete sequence of the *MwCP* 5'-end.
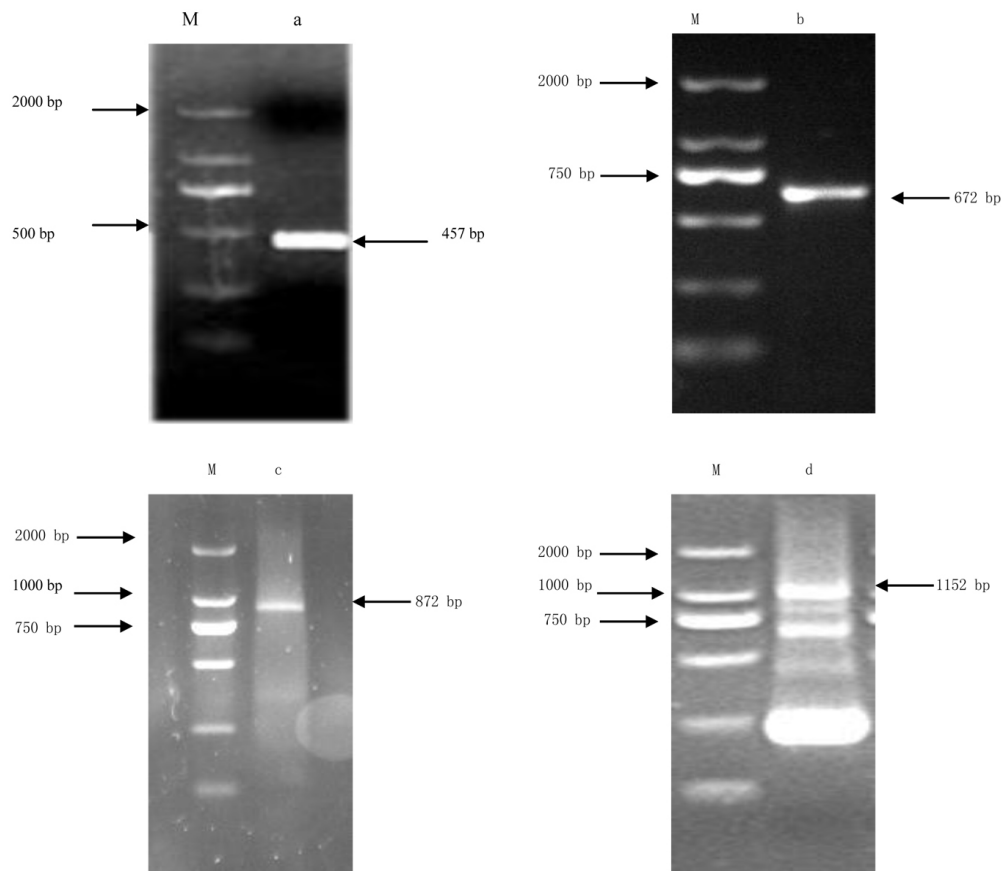
Figure 1. Race amplification results. **a.** 3'-RACE amplification products were detected by agarose gel electrophoresis, and a ~500-bp band was observed (457 bp after sequencing). **b.** A ~700-bp fragment was obtained from PCR, and the precise length was 672 bp according to sequencing results. **c.** Using the sequence obtained from the 5'-RACE analysis, an 827-bp fragment was observed following the second nested PCR analysis. **d.** The *MwCP* ORF was identified using ORF Finder, and the CP-FSac2 and CP-RXho2 primers were designed for ORF amplification. A 1200-bp fragment was obtained from PCR (1152 bp after sequencing). *Lane M* = molecular marker.

Acquisition of *MwCP* and the ORF

We assembled the sequences obtained from 3'-RACE and 5'-RACE using the DNAMAN program, and we obtained a 1481-bp cDNA sequence, which was named *MwCP* because it was a CP gene found in *A. mongolicum*. The cloned cDNA sequence was submitted to GenBank under the accession No. KJ534636. We found the ORF of *MwCP* using the ORF Finder program, and we designed the CP-FSac2 and CP-RXho2 primers used to amplify the ORF. A 1152-bp band was obtained from PCR (Figure 1d), and the complete length of the ORF was 1134 bp after removing the restriction enzyme sites.

Bioinformatic analysis

Search for the MwCP ORF

The full-length *MwCP* cDNA sequence was 1473 bp. The ORF search results indicated that the cDNA sequence contained a 1134-bp ORF, which encoded 377 amino acids (Figure 2) and two untranslated regions (26 bp at the 5'-end and 313 bp at the 3'-end). In addition, there was a poly-A tail at the 3'-end. Based on the gene length, the ORF, and the number of encoded amino acids, the *MwCP* gene was essentially consistent with the *CsCysP* gene isolated from citrus (Ma et al., 2014), which indicated that *MwCP* was a CP gene.

Structure prediction of the MwCP protein

The results of the conserved domain analysis of the MwCP protein indicated that the protein contained one transmembrane region, one Inhibitor_I29 domain, and one Pept_C1 domain (Figure 3). The results were further confirmed by NCBI and the SMART software, which reported identical results. In short, the protein was a CP that could be categorized as a member of the Papain C1 family.

The secondary structure of the coded protein mainly consisted of α-helices and random coils (Figure 4A). The three-dimensional structure prediction of the coded protein indicated that the space structure mainly contained α-helices, β-sheets, and β-turns (Figure 4B), which was in agreement with a study of CP in *Brassica napus* (Zhang et al., 2008).

Physical and chemical property analyses of the MwCP protein

Analyses of the physical and chemical properties of the MwCP protein showed that the molecular weight was 41.01 kDa, which was consistent with that of a CP in wheat and the protein encoded by the *HbCP2* and *HbCP3* genes in the Para rubber tree (*Hevea brasiliensis*) (Liang et al., 2014; Savvateeva et al., 2015). Furthermore, the pI was 5.80, the formula was $C_{1803}H_{2793}N_{507}O_{552}S_{19}$, and the instability index was 36.03, which meant that the protein was stable based on the standard (proteins with values <40 are stable). Among the 20 kinds of amino acids that composed the protein, glycine (Gly) accounted for the highest proportion (up to 10.1%), and tryptophan (Trp) exhibited the lowest proportion (1.3%). The total number of negatively charged residues (Asp + Glu) was 48, and the total number of positively charged residues (Arg + Lys) was 40. The aliphatic index was 73.53, and the grand average of hydropathicity (GRAVY) was -0.325.

```
1      ACCCAAACAAGCAAGAAAATTCCCCAATGGATCATCGCCTCGTGGCCCCCCTCCTCGTCC
1                                       M  D  H  R  L  V  A  P  L  L  V
61     TCCTGGGCCTCCTCCTCTCCCCGGCGCCGGCAGCGGCGGCCGCGGGGGACGAGGAACCCC
12     L  L  G  L  L  L  S  P  A  P  A  A  A  A  G  D  E  P
121    TGATCCGGCAGGTCGTCGGGGGCGCCGACCCCCTCGACAACGACCTGGAGCTCGACTCGC
32     L  I  R  Q  V  V  G  A  D  P  L  D  N  D  L  E  L  D  S
181    AGTTCGTCGGCTTCGTGCAGCGGTTCGGGAAGACGTACAGGGACGCGGAGGAGCACGCGC
52     Q  F  V  G  F  V  Q  R  F  G  K  T  Y  R  D  A  E  E  H  A
241    ACCGGCTCTCCGTGTTCAAGGCCAACCTCCGCCGCGCGCGCCGGCACCAGATGCTCGACC
72     H  R  L  S  V  F  K  A  N  L  R  R  A  R  R  H  Q  M  L  D
301    CGTCCGCGGAGCACGGGGTCACCAAGTTCTCCGACCTCACCCCGGCCGAGTTCCGCCGGA
92     P  S  A  E  H  G  V  T  K  F  S  D  L  T  P  A  E  F  R  R
361    CCTTCCTGGGGCTCAAGACGACCCGGCGATCGTTCCTGCGGGAGATGGCCGGGTCGGCGC
112    T  F  L  G  L  K  T  T  R  R  S  F  L  R  E  M  A  G  S  A
421    ACGACGCGCCCGTCCTCCCCACCGACGGCCTCCCCGAGGACTTCGACTGGAGGGACCACG
132    H  D  A  P  V  L  P  T  D  G  L  P  E  D  F  D  W  R  D  H
481    GCGCCGTCGGCCCCGTCAAGAACCAGGGGTTCGTGCGGGTCGTGCTGGTCGTTCAGCGCCT
152    G  A  V  G  P  V  K  N  Q  G  S  C  G  S  C  W  S  F  S  A
541    CCGGGGCGTTGGAGGGGGCCAACTACCTGGCCACGGGCAAGATGGAGGTGCTCTCCGAGC
172    S  G  A  L  E  G  A  N  Y  L  A  T  G  K  M  E  V  L  S  E
601    AGCAGATGGTCGACTGCGACCATGAGTGCGACCCAGCAGAACCTGATTCATGCGATGCTG
192    Q  Q  M  V  D  C  D  H  E  C  D  P  A  E  P  D  S  C  D  A
661    GATGCAACGGTGGGTTGATGACCTCAGCCTTCAGCTATCTGTTGAAATCTGGTGGCCTTG
212    G  C  N  G  G  L  M  T  S  A  F  S  Y  L  L  K  S  G  G  L
721    AGAGAGAAAAGGATTACCCTTACACCGGGAAGGATGGTACCTGCAAATTTGAGAAGTCCA
232    E  R  E  K  D  Y  P  Y  T  G  K  D  G  T  C  K  F  E  K  S
781    AGATTGCTGCTTCAGTTCAGAACTTCAGCGTTGTCGCTGTTGATGAAGAACAGATTGCTG
252    K  I  A  A  S  V  Q  N  F  S  V  V  A  V  D  E  E  Q  I  A
841    CTAACCTTGTCAAATATGGACCGCTGGCCATCGGTATCAACGCCGCTTACATGCAGACGT
272    A  N  L  V  K  Y  G  P  L  A  I  G  I  N  A  A  Y  M  Q  T
901    ACATCGGCGGAGTGTCATGCCCATACATCTGCGGCAGGCACCTCGACCACGGCGTCCTCC
292    Y  I  G  G  V  S  C  P  Y  I  C  G  R  H  L  D  H  G  V  L
961    TCGTAGGCTACGGGGCGTCCGGCTTCGCACCTTCCCGCTTCAAGGAGAAGCCCTACTGGA
312    L  V  G  Y  G  A  S  G  F  A  P  S  R  F  K  E  K  P  Y  W
1021   TCATCAAGAACTCATGGGGCGAGAACTGGGGGGACAAGGGGTACTACAAGATCTGCAGGG
332    I  I  K  N  S  W  G  E  N  W  G  D  K  G  Y  Y  K  I  C  R
1081   GCTCCAACGTGCGCAACAAGTGCGGCGTCGACAACATGGTCTCCACCGTGTCCGCCACGC
352    G  S  N  V  R  N  K  C  G  V  D  N  M  V  S  T  V  S  A  T
1141   ATGCCTCCAAGGAGGAGTAGGCTCTGGTCTGGTCTGATCTGATCTGATCGGCGGCCCTCC
372    H  A  S  K  E  E  *
1201   TGGAGTCGATCTTGGTTCCGGTGTATCGCTATTCGCTAGAAAGAAACCCTTAATACGTAG
1261   TGGTCGGCTTAGGCTCCATCTTGCCGCTGCGGTACGAGCAGCAGCGGAGATGCGAGCCGC
1321   AACAGAATGCTTGCTGTATAACTTATGCATTTGCTCTATTTGCTACGCCATGCATGTTAT
1381   GTTTGCCACACGCTATTTGGATGTGGCTAAAGAACTTCTGAATAATTATGTACATAATAT
1441   CGATTGCTTCAAAAAAAAAAAAAAAAAAAAAAAAA
```

Figure 2. Full-length (1473 bp) *MwCP* cDNA sequence. The ORF search was performed using ORF Finder (NCBI), and the results showed that the cDNA sequence contained a 1134-bp ORF. The ORF encoded 377 amino acids and two untranslated regions (a 26-bp 5'-end and a 313-bp 3'-end). In addition, there was a poly-A tail at the 3'-end.
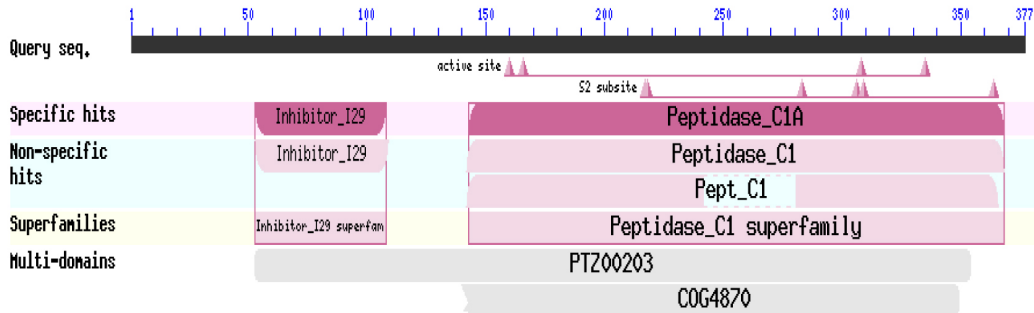
Figure 3. MwCP conserved domains. The conserved domain analysis of the MwCP protein was performed using the SMART software and the NCBI Conserved Domain Database. The results indicated that MwCP contained one transmembrane region (23 amino acid residues) between the 5th and the 27th positions, one Inhibitor_I29 domain (57 amino acid residues) between the 53rd position and the 109th positions, and one Pept_C1 domain (228 amino acid residues) between the 142nd and the 369th positions.
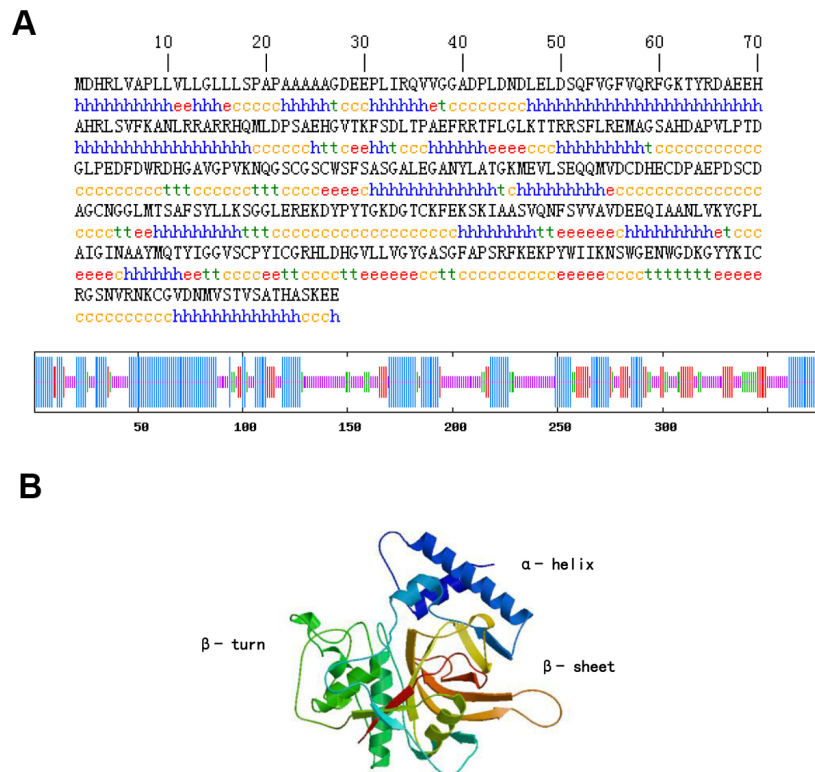

Figure 4. MwCP secondary structure. **A.** Results of the MwCP secondary structure analysis (predicted using the SOPMA software) indicated that 152-amino acid residues (40.32%) formed α-helices, 48-amino acid residues (12.73%) were for extended strands, 36-amino acid residues (9.55%) were β-turns, and 141-amino acid residues (37.40%) made up random coils. **B.** Results of the three-dimensional structure of the coded protein (predicted by SWISS-MODEL) indicated that the spatial structure mainly contained α-helices, β-sheets, and β-turns.

In addition, a Kyte and Doolittle hydropathic plot that was generated using the ProtScale program indicated that the protein was hydrophilic (Figure 5).
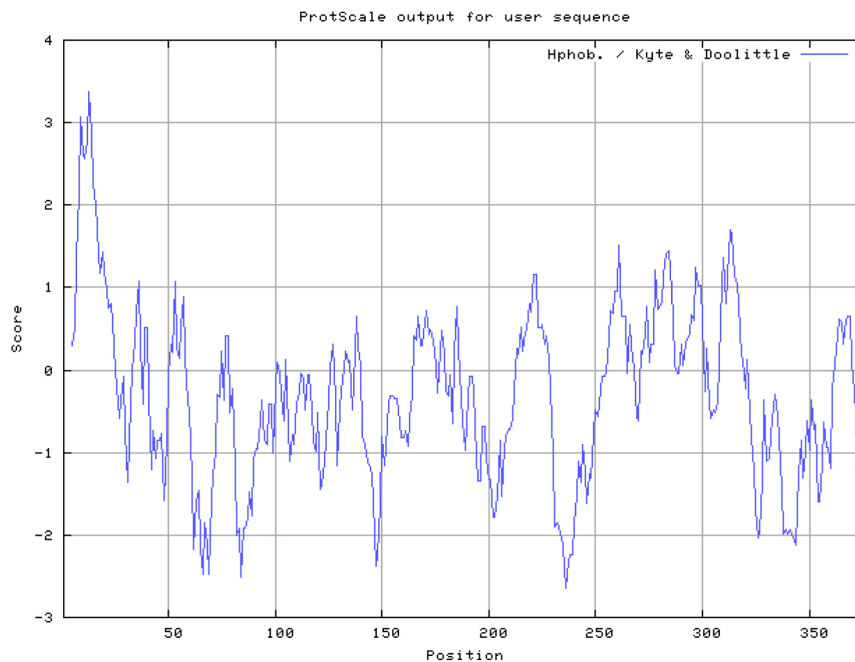
Figure 5. MwCP physical and chemical properties. Results of the MwCP physical and chemical property analysis (conducted using the ProtParam software) showed that the maximum hydrophobic value of 3.378 was located in the 13th leucine (Leu), and the maximum hydrophilic value of 2.644 was located in the 236th aspartic acid (Asp). There was a typical hydrophobic region in the previous 24 amino acids of the N-terminal, but most of the amino acids were hydrophilic.

Functional analysis of the MwCP protein

The MwCP protein was found to be extracellular based on the results of the PSORT and Wolf PSORT analyses. The results of the SignalP analysis showed that the MwCP protein had obvious characteristics of an N-terminal signal peptide (Figure 6), and the TargetP and iPSORT results further supported the conclusion that there was an N-terminal signal peptide. Several CPs contain N-terminal signal peptides, such as the CPs triticain-α and gliadain in wheat and OsCP2 in rice (Kiyosaki et al., 2007; Yan et al., 2010; Savvateeva et al., 2015). The results of the TMHMM analysis indicated that the N-terminal of the MwCP protein had a transmembrane region from the inside to the outside between the 5th and 27th positions. The MwCP was predicted to be a secretory protein. The result was consistent with several researches, such as the research about CP produced by suspension of rice cell (Kim et al., 2008) and the research about CP (*HbCP2*) from *H. brasiliensis* (Zou et al., 2014).

The results of the PROSITE program analysis showed that the Pept_C1 domain contained three CP active sites: THIOL_PROTEASE_CYS (160 to 171: QGSCGSCWSFSA) eukaryotic thiol (cysteine) protease cysteine active site, THIOL_PROTEASE_HIS (306 to 316: LDHGVLLVGY)

eukaryotic thiol (cysteine) protease histidine active site, and THIOL_PROTEASE_ASN (330 to 349: YWIIKNSWGENWGDKGYYKI) eukaryotic thiol (cysteine) protease asparagine active site. Among the above sequences, the 166th C, the 308th H, and the 335th N were active sites that determined CP catalytic activity. These active sites are common characteristics of the Papain family-catalytic triad ($Cys_{166}His_{308}Asn_{335}$).
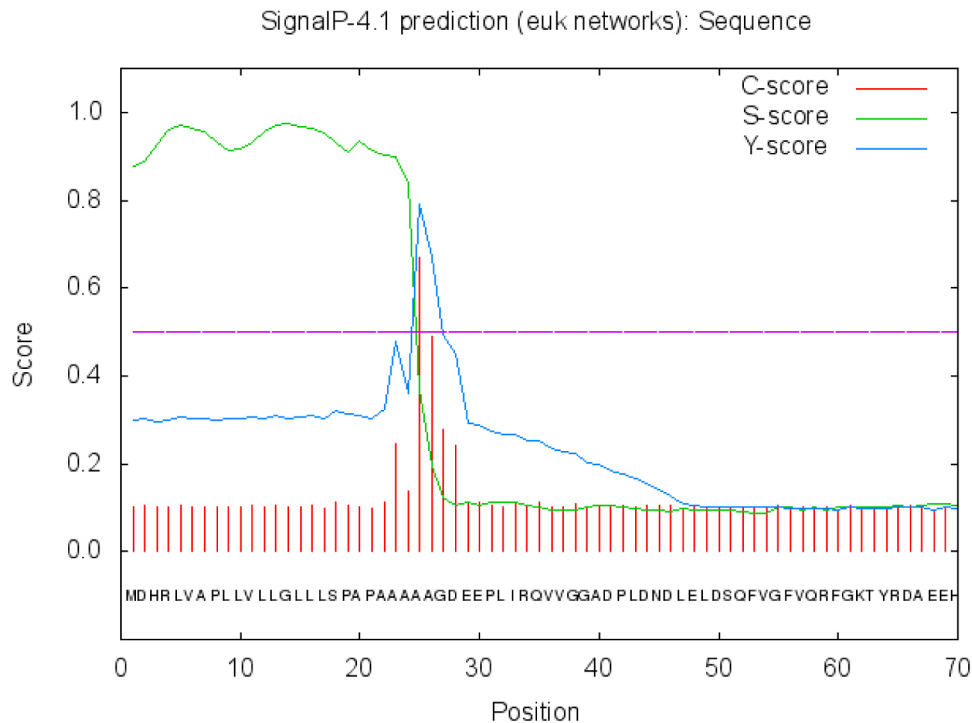
Figure 6. MwCP functional results. The results of the SignalP analysis showed that MwCP had obvious characteristics of an N-terminal peptide signal, and the possible cleavage site was between the 24th and 25th positions.

MwCP protein homology analysis and phylogenetic tree construction

The results of the BLASTp analysis of the MwCP amino acid sequence showed that the protein shared >90% similarity with homologs found in *T. aestivum* (AAW21813.1), *H. vulgare* subsp *vulgare* (CAD66657.1), *Aegilops tauschii* (EMT10192.1), and *Brachypodium distachyon* (XP_0035 74963.1).

A multiple-sequence alignment of the MwCP protein and CPs from *T. aestivum* (AAW21813.1), *H. vulgare* subsp *vulgare* (CAD66657.1), *A. tauschii* (EMT10192.1), and *B. distachyon* (XP_003574963.1) indicated that the MwCP protein was highly homologous (96.60%) to the CPs of other plants (Figure 7). The sequence alignment of the plant CPs revealed that while CP had a relatively high conservative C-terminal region and a middle region containing CP active sites, there was a low degree of diversity in the N-terminal region, including the signal peptide.

Figure 7. MwCP homology results. The multiple-sequence alignment (using DNAMAN) of MwCP from *Agropyron mongolicum* Keng and CPs of *Triticum aestivum* (AAW21813.1), *Hordeum vulgare* subsp *vulgare* (CAD66657.1), *Aegilops tauschii* (EMT10192.1), and *Brachypodium distachyon* (XP_003574963.1) indicated that MwCP was highly homologous to CPs of other plants (96.60%).

　　　　Cluster analysis of *MwCP* and 16 known CP genes from other plants at the amino acid level showed that *MwCP* was closest to the CP of *T. aestivum*, and both species belong to Triticeae. Furthermore, there was a relatively intermediate relationship between *MwCP* and the CPs of *A. tauschii*, *H. vulgare* subsp *vulgare*, and *B. distachyon*. Among the CPs, the most distant relationship existed between *MwCP* and the CPs of *O. sativa* (*japonica* group), *Setaria italica*, *Zea mays*, and *Sorghum bicolor*. However, the evolutionary relationships suggested that *MwCP* was also quite distant from dicotyledonous plants such as *Glycine max*, *H. brasiliensis*, *Arabidopsis thaliana*, *Ricinus communis,* and *Phaseolus vulgaris* (Figure 8). In conclusion, the results suggest that gramineous monocotyledonous CP genes were very close relatives, and *MwCP* was distant from other species, especially dicotyledonous plants. The above clustering results were consistent with biological classifications, which suggested that the CP genes of certain species were evolutionarily conserved. These results were largely consistent with those found in a relevant study of *T. aestivum* (Kim et al., 2008), which indicated that *MwCP* was involved in drought-resistance in plants.
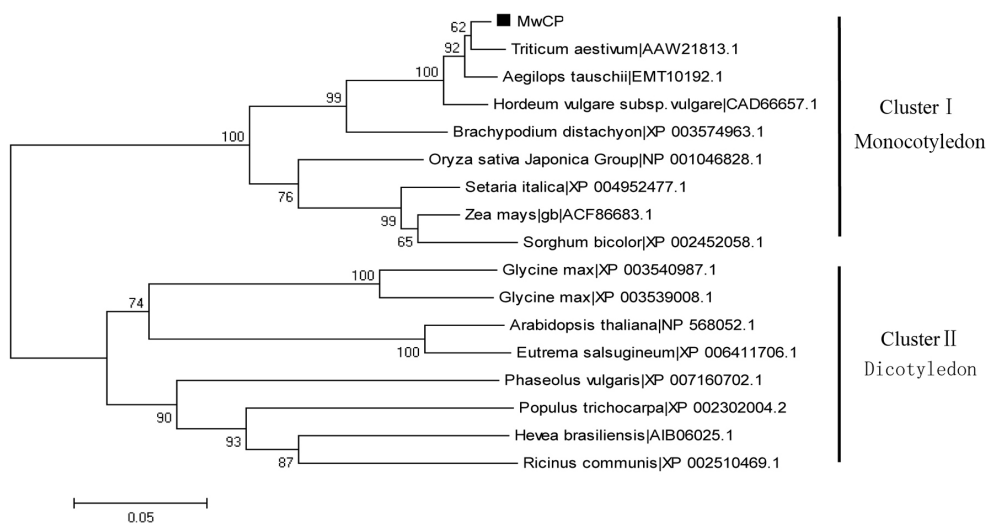

Figure 8. MwCP cluster analysis results. The results of the cluster analysis among amino acid sequences encoded by *MwCP* and 16 other known plants (sequences downloaded from GenBank and analysis conducted using BLAST and MEGA software) indicated that 17 kinds of plant CPs were classified into two large groups: group A (cluster I) was monocotyledonous plants and nine kinds of gramineous plants; group B (cluster II) was dicotyledonous plants.

Expression analysis of *MwCP*

　　　　Temporal specificity (Figure 9) and tissue specificity (Figure 10) were analyzed using semi-quantitative RT-PCR. When subjected to drought stress, the expression products appeared during the 4th hour, which was in accordance with the expression of the *TaCP* gene isolated from common wheat by Zang et al. (2010). Moreover, the expression levels of both *MwCP* and *TaCP* gradually increased with time. There were also obvious changes at different times, which indicated that *MwCP* was activated during drought stress to gradually produce CP that was used to degrade inactive proteins. Sun et al. (2004) found that leaves were most sensitive to drought, and they

suffered the most damage. Parkash et al. (2015) also found that the expression levels of CP genes increased during leaf senescence. The leaves exhibited the highest amount of expression products, which suggested that *MwCP* was involved in drought-resistant physiological processes in the plant.
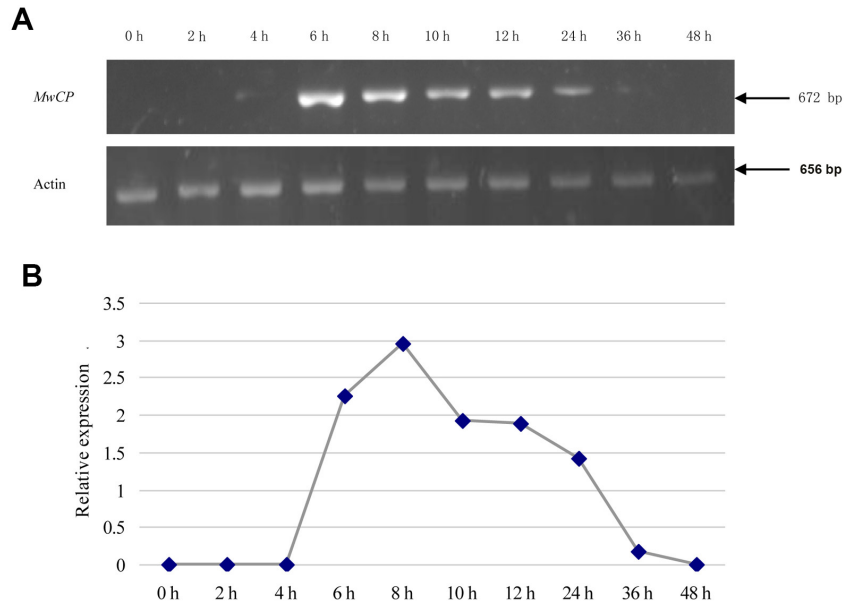

Figure 9. MwCP expression results. Temporal specificity of *MwCP* was analyzed using semi-quantitative RT-PCR. **A.** During the 48-h drought stress, the expression products appeared during the 4th hour, and they peaked at the 6th hour. They gradually declined later, and reached the initial level during the 36th hour. The expression then ceased, and no expression could be detected during the 48th hour. **B.** Relative expression of *MwCP* during different drought stress periods.
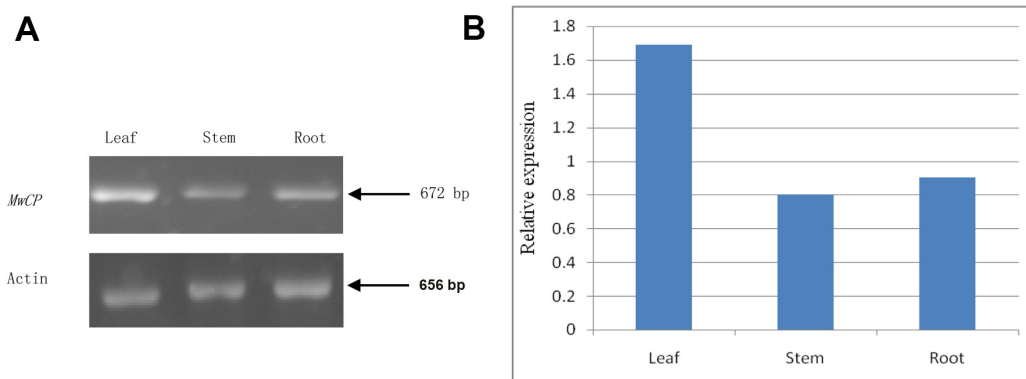

Figure 10. MwCP tissue specificity results. *MwCP* tissue specificity (**A**) was analyzed using semi-quantitative RT-PCR. Different tissues were tested, including leaves, roots, and stems, and the amount of the expression products was highest in the leaf, intermediate in root, and lowest in the stem. **B.** Relative expression of *MwCP* in different tissues.

DISCUSSION

CPs are important proteases that are involved in plant drought resistance. CP members were divided into many families such as papain (C1), legumain (C13), caspase (C14), calpain (C2), and ubiquitin-like CP (Basset et al., 2002; Vierstra, 2003). CP genes were previously extracted via the cloning of plants like *T. vulgare* (Zang et al., 2010), *Gossypium* spp (Jiang and Zhang, 2004), and *H. brasiliensis* (Zhu, 2007). However, this experiment represents the first time that a CP gene was obtained from the cloning of *A. mongolicum*. The 1473-bp *MwCP* cDNA contained a 1134-bp ORF, and it encoded a 377-amino acid CP (Figure 2). The protein encoded by *MwCP* exhibited a secondary structure that was mainly α-helices and random coils (Figure 4). The protein was also a secretory stable hydrophilic protein with a molecular weight of 41.01 kDa, which indicated that it played a part in drought stress outside of the cells. The encoded protein had an N-terminal signal peptide (Figure 6) and highly conserved Inhibitor_I29 and Pept_C1 domains (Figure 3). The Pept_C1 domain contained a catalytic triad (CysHisAsn) associated with CPs. Results of analyses of *MwCP* cDNA, the structural and physicochemical properties of the encoded protein, etc., were in accordance with those of the previously studied papain C1 family (Jones and Mullet, 1995; Grudkowska and Zagdańska, 2004). Therefore, the coded protein was a new member of papain C1 family, and it was likely involved in drought resistance in plants (Battelli et al., 2014).

Semi-quantitative RT-PCR analyses revealed that the *MwCP* expression levels gradually increased with drought stress over time, and there were obvious changes at different times (Figure 9), which was similar to the results of a relevant study of CPs in *T. aestivum* (Zang et al., 2010). In addition, the *MwCP* expression levels were obviously upregulated in the leaf (Figure 10), which indicated that the CP mainly degraded inactive proteins in the leaves (Yan et al., 2005; Zang et al., 2010). This particular characteristic was closely associated with *A. mongolicum* drought resistance.

Homology analyses showed that the encoded protein had high homology (96.60%) to *T. aestivum*, *H. vulgare* subsp *vulgare*, *A. tauschii*, and *B. distachyon* (Figure 7). This result indicated that *MwCP* had relatively conserved characteristics during its evolution. Phylogenetic analysis (Figure 8) indicated that the CP genes of Gramineae were very close relatives, and *MwCP* was distant from other species, especially dicotyledonous plants. This result further suggested the conservation of CP genes in certain species during the evolutionary process, and this characteristic could be used in genetic engineering to improve the adverse resistance of plants. Zang et al. (2010) previously cloned CP genes from *T. aestivum* and transferred them to *Arabidopsis thaliana,* which resulted in enhanced drought resistance of *A. thaliana*. Both *A. mongolicum* and *T. aestivum* belong to Triticeae (Gramineae), and *A. mongolicum* has high drought resistance. Therefore, *MwCP* may be expressed in other plants to enhance their drought resistance, but further study is needed to confirm this hypothesis.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Aberlenc-Bertossi F, Chabrillange N, Duval Y and Tregear J (2008). Contrasting globulin and cysteine proteinase gene expression patterns reveal fundamental developmental differences between zygotic and somatic embryos of oil palm. *Tree Physiol.* 28: 1157-1167.http://dx.doi.org/10.1093/treephys/28.8.1157

Basset G, Raymond P, Malek L and Brouquisse R (2002). Changes in the expression and the enzymic properties of the 20S proteasome in sugar-starved maize roots. Evidence for an *in vivo* oxidation of the proteasome. *Plant Physiol.* 128: 1149-1162.http://dx.doi.org/10.1104/pp.010612

Battelli R, Lombardi L, Picciarelli P, Lorenzi R, et al. (2014). Expression and localisation of a senescence-associated KDEL-cysteine protease from *Lilium longiflorum* tepals. *Plant Sci.* 214: 38-46.http://dx.doi.org/10.1016/j.plantsci.2013.09.011

Che YH and Yang YP (2010). Genetic diversity of *Agropyron mongolicum* Keng indigenous to northern China. 2010 First International Conference on Cellular, Molecular Biology, Biophysics and Bioengineering (*CMBB*), 45-47.

Chen HJ, Huang DJ, Hou WC, Liu JS, et al. (2006). Molecular cloning and characterization of a granulin-containing cysteine protease SPCP3 from sweet potato (*Ipomoea batatas*) senescent leaves. *J. Plant Physiol.* 163: 863-876. http://dx.doi.org/10.1016/j.jplph.2005.08.008

Cui KM (2000). Mechanisms of programmed cell death and its relation to development in plants. *Chinese Bull. Bot.* 17: 97-107.

Griffiths CM, Hosken SE, Oliver D, Chojecki J, et al. (1997). Sequencing, expression pattern and RFLP mapping of a senescence-enhanced cDNA from *Zea mays* with high homology to oryzain gamma and aleurain. *Plant Mol. Biol.* 34: 815-821.http://dx.doi.org/10.1023/A:1005896713830

Grudkowska M and Zagdańska B (2004). Multifunctional role of plant cysteine proteinases. *Acta Biochim. Pol.* 51: 609-624.

Jiang JX and Zhang TZ (2004). Cloning and sequence analysis of a full-length cysteine proteinase cDNA in *Gossypium hirsutum* L. *Acta Agron. Sin.* 30: 512-515.

Jones JT and Mullet JE (1995). A salt- and dehydration-inducible pea gene, Cyp15a, encodes a cell-wall protein with sequence similarity to cysteine proteases. *Plant Mol. Biol.* 28: 1055-1065.http://dx.doi.org/10.1007/BF00032666

Kim NS, Kim TG, Kim OH, Ko EM, et al. (2008). Improvement of recombinant hGM-CSF production by suppression of cysteine proteinase gene expression using RNA interference in a transgenic rice culture. *Plant Mol. Biol.* 68: 263-275. http://dx.doi.org/10.1007/s11103-008-9367-8

Kiyosaki T, Matsumoto I, Asakura T, Funaki J, et al. (2007). Gliadain, a gibberellin-inducible cysteine proteinase occurring in germinating seeds of wheat, *Triticum aestivum* L., specifically digests gliadin and is regulated by intrinsic cystatins. *FEBS J.* 274: 1908-1917.

Li LH and Dong YC (1993). Progress in studies of *Agropyron* Gaertn. *Hereditas* 43: 45-52.http://dx.doi.org/10.1159/000154113

Li YX (2008). The function of a cysteine proteinase gene, *Asnodf32*, in *Astragalus sinicus* nodule senescence. Doctoral thesis. Huazhong Agriculture University.

Liang QF, Xin LS, Long XY and ChaoRong T (2014). Cloning and expression analysis of cysteine proteinase genes (*HbCP2* and *HbCP3*) from para rubber tree (*Hevea brasiliensis*). *J. Agric. Biotechnol.* 22: 690-702.

Ma YY, Zhang J, Chen J, et al. (2014). Isolation, subcellular localization and expression analysis of a citrus cysteine protease gene, Cs Cys P. *Acta Hort. Sin.* 41: 621-630.

Mohan S, Ma PW, Pechan T, Bassford ER, et al. (2006). Degradation of the *S. frugiperda* peritrophic matrix by an inducible maize cysteine protease. *J. Insect Physiol.* 52: 21-28.http://dx.doi.org/10.1016/j.jinsphys.2005.08.011

Nassar NMA, Abreu LFA, Teodoro DA and Graciano-Ribeiro D (2010). Drought tolerant stem anatomy characteristics in *Manihot esculenta* (Euphorbiaceae) and a wild relative. *Genet. Mol. Res.* 9: 1023-1031. http://dx.doi.org/10.4238/vol9-2gmr800

Parkash J, Kashyap S, Kirti S, Singh AK, et al. (2015). Cathepsin B cysteine protease gene is upregulated during leaf senescence and exhibits differential expression behaviour in response to phytohormones in *Picrorhiza kurrooa* Royle ex Benth. *Plant Gene* 3: 11-19.http://dx.doi.org/10.1016/j.plgene.2015.07.001

Reis RR, da Cunha BA, Martins PK, Martins MTB, et al. (2014). Induced over-expression of *AtDREB2A CA* improves drought tolerance in sugarcane. *Plant Sci.* 221-222: 59-68.http://dx.doi.org/10.1016/j.plantsci.2014.02.003 PubMed

Savvateeva LV, Gorokhovets NV, Makarov VA, Serebryakova MV, et al. (2015). Glutenase and collagenase activities of wheat cysteine protease Triticain-α: feasibility for enzymatic therapy assays. *Int. J. Biochem. Cell Biol.* 62: 115-124. http://dx.doi.org/10.1016/j.biocel.2015.03.001

Stolf-Moreira R, Medri ME, Neumaier N, Lemos NG, et al. (2010). Soybean physiology and gene expression during drought. *Genet. Mol. Res.* 9: 1946-1956.http://dx.doi.org/10.4238/vol9-4gmr851

Sun MX, Zu ZL and Xu JN (2004). Research progress on the impact of drought on plant. *J. Anhui Agric. Sci.* 32: 365-367.

Vierstra RD (2003). The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. *Trends Plant Sci.* 8: 135-142.http://dx.doi.org/10.1016/S1360-1385(03)00014-1

Wang HQ, Deng XY, Long QW, Peng H, et al. (2012). Cysteine protease of plant potential target of herbicide. *Agrochemicals* 51: 405-408.

Yamaguchi-Shinozaki K and Shinozaki K (1993). *Arabidopsis* DNA encoding two desiccation-responsive *rd29* genes. *Plant Physiol.* 101: 1119-1120.http://dx.doi.org/10.1104/pp.101.3.1119

Yan LF, Yang QC, Han JG and Lui Z (2005). Summary of cysteine protease in plants. *Acta Pratacultural Sci.* 14: 11-19.

Yan XR, Zhang DS, Liang WQ, Wang Y, et al. (2010). Characterization, prokaryotic expression and purification of a rice cysteine protease OsCP2. *J. Shanghai Jiaotong Univ. (Agric. Sci.*) 28: 140-146.

Yuan LX, Sun Y, Yang YJ and Zhao HF (2008). The research on mechanism of programmed cell death. *J. Shanxi Agric. Sci.* 36: 15-17.

Zang QW, Wang CX, Li XY, Guo ZA, et al. (2010). Isolation and characterization of a gene encoding a polyethylene glycol-induced cysteine protease in common wheat. *J. Biosci.* 35: 379-388.http://dx.doi.org/10.1007/s12038-010-0043-1

Zhang HY, Xue H, Ma XR, Zhao Y, et al. (2008). Full-length cDNA cloning and tissue specified expression analysis of a cysteine protease in *Brassica napus* L. *Chin. J. Appl. Environ. Biol.* 14: 172-176.

Zhang Z, Shi WG and Lu XP (2004). Advances in study on *Agropyron mongolicum* Keng. *Pratacultural Sci.* 7: 459-465.

Zhao Y (2009). Isolation, expression and construction RNAi vector of drought-resistant related genes in Mongolian wheatgrass. Inner Mongolia Agriculture University. Doctoral thesis.

Zhao Y, Yun JF, Shi FM and Gao CP (2008). Isolation and characterization of a *Lea3* gene fragment from *Agropyron mongolicum* Keng. *Acta Agric. Bor. Sin.* 23: 64-67.

Zhu JH (2007). Cloning and expression analysis on of cysteine protease gene *HbCP1* from *Hevea brasiliensis*. South China University of Tropical Agriculture. Master's thesis.

Zhu JH, Li HL and Peng SQ (2009). Expression analysis of the cysteine protease gene (*HbCP1*) in *E. coli. Chinese J. Trop. Crop.* 30: 667-671.

Zou Z, Liu JT, Zhuang ML and Yang LF (2014). Molecular cloning and expression analysis of a cysteine proteinase gene (HbCP2) from *Hevea brasiliensis. J. Southwest Forestry Univ.* 34: 26-30.