



Cloning and characterization of endo- β -1,4-glucanase genes in the common wheat line three pistils

Zaijun Yang, Zhengsong Peng, Shuhong Wei and Yan Yu

Key Laboratory of Southwest China Wildlife Resources Conservation, China West Normal University, Nanchong City, Sichuan, People's Republic of China.

Abstract

In this work, we report the cloning and characterization of endo- β -1,4-glucanase (EGase) genes (TaEG) in the common wheat line three pistils. Three TaEG homoeologous genes (TaEG-4A, TaEG-4B and TaEG-4D) were isolated and found to be located on chromosomes 4AL, 4BS and 4DS, respectively. The three genes showed high conservation of their coding nucleotide sequences and 3 untranslated region. The putative TaEG protein had a molecular mass of 69 kDa, a theoretical pI of 9.39 and a transmembrane domain of 74-96 amino acids in the N-terminus that anchored the protein to the membrane. The genome sequences of TaEG-4A, TaEG-4B and TaEG-4D contained six exons and five introns. All of the introns, except for intron IV, varied in length and sequence composition. Phylogenetic analysis revealed that TaEG was most closely related to rice (*Oryza sativa*) OsGLU1. The TaEG transcript levels increased significantly during the subsidiary pistil primordium differentiation phase (spike size ~7-10 mm) in Chuanmai 28 TP (CM28TP). These data provide a basis for future research into the function of TaEG and offer insights into the molecular mechanism of the three pistils mutation in wheat.

Keywords: cloning, endo- β -1,4-glucanase, three pistils line, wheat.

Received: April 26, 2013, Accepted: June 3, 2013.

Introduction

Endo- β -1,4-glucanases (EGases) are enzymes that hydrolyze polysaccharides containing a 1,4- β -glucan backbone and are produced by prokaryotes and eukaryotes (Henrissat *et al.*, 1989; Beguin, 1990; Watanabe *et al.*, 1998; Byrne *et al.*, 1999; Rosso *et al.*, 1999). These enzymes have also been classified as cellulases because of their central role in cellulose degradation and their potential to modify cellulose-containing materials (Bisaria and Mishra, 1989; Beguin, 1990; Beguin and Aubert, 1994). Sequence analysis has shown that all cloned plant EGases can be classified into two subfamilies (Nicol *et al.*, 1998). One of these subfamilies consists of soluble secreted EGases such as avocado Cel1, pepper CX1 and tomato Cel2 that are expressed specifically during fruit ripening (Christoffersen *et al.*, 1984; Lashbrook *et al.*, 1994; Ferrarese *et al.*, 1995), while others such as bean BAC1, soybean SAC1 and elder JET1, are associated with abscission (Tucker *et al.*, 1988; Kemmerer and Tucker, 1994; Taylor *et al.*, 1994). The other subfamily consists of membrane-anchored proteins located in the plasma membrane that also contribute to nor-

mal cellulose formation (Brummell *et al.*, 1997a; Nicol *et al.*, 1998; Del Campillo, 1999; Molhoj *et al.*, 2001). This subfamily includes the *Arabidopsis* KOR gene and the *O. sativa* OsGLU1 gene that are involved in cell elongation (Nicol *et al.*, 1998; Lane *et al.*, 2001; Zhou *et al.*, 2006).

The common wheat (*Triticum aestivum* L.) line three pistils (TP), which was selected by Peng (2003) from the "trigrain" wheat cultivar, is a valuable mutant for improving wheat yield. Since the TP mutation has normal spike morphology but produces three pistils per floret it has potential for increased grain number per spike. In a previous study, a sequence from an expressed sequence tag (EST) was initially identified in a TP mutation in wheat using an annealing control primer system (ACP) (Yang *et al.*, 2011). Blast searches of this sequence against the GenBank database revealed that it was homologous to the EGase gene. To our knowledge, the role of EGase genes in wheat development has not yet been reported.

As part of an investigation into the function of the wheat EGase gene (TaEG) in pistil development in this work we have cloned, characterized and phylogenetically analyzed the TaEG gene in the common wheat line three pistil. We also examined the expression patterns of this gene in different developmental stages of young spikes.

Materials and Methods

Plant material

Chuanmai 28 TP (CM28TP), a near-isogenic line of the common wheat line Chuanmai 28 carrying the *Pis1* gene derived from the TP mutant (Yang *et al.*, 2012), and the recurrent parent *T. aestivum* cv. Chuanmai 28 (CM28) were used in this study. The near-isogenic line and the recurrent parent were cultivated in the field at the China West Normal University in Nanchong, China. Young spikes at the reproductive stage were collected and examined with an Olympus SZX9 stereomicroscope to accurately determine their developmental stage. Subsidiary pistil spikes in the primordial stage of differentiation (spike size ~7-10 mm) were used as a source of RNA and DNA for cloning the cDNA and sequencing the TaEG gene. Spikes of CM28TP and CM28 at various developmental stages (spikes 2-5, 5-7, 7-10, 10-13, 13-15, 15-17 and 17-20 mm in length) were selected for real-time PCR. In addition, the nullisomic-tetrasomic (NT) lines and ditelosomic (Dt) lines Dt1AL, Dt1AS, Dt2AS, Dt3AL, Dt3AS, Dt4AL, N4AT4B, Dt5AL, N5AT5B, Dt6AL, Dt6AS, Dt7AL, Dt7AS, Dt1BL, Dt1BS, Dt2BL, N2BT2A, Dt3BL, Dt3BS, Dt4BL, Dt4BS, Dt5BL, N5BT5D, Dt6BS, N6BT6D, Dt7BL, Dt7BS, Dt1DL, Dt1DS, Dt2DL, Dt2DS, Dt3DL, Dt3DS, Dt4DL, Dt4DS, Dt5DL, N5DT5A, Dt6DL, Dt6DS, Dt7DL and Dt7DS derived from CS (Sears, 1954; Endo and Gill, 1996) were used for the chromosomal mapping of TaEG.

DNA and RNA isolation

Genomic DNA was isolated from young spikes of CM28TP and CM28 and fresh leaves of the NT and Dt lines, as described by Porebski *et al.* (1997). The DNA was dissolved in Tris-EDTA (TE) buffer and stored at -20 °C. Total RNA was isolated from young spikes as described by Manickavelu *et al.* (2007). The RNA was dissolved in RNase-free double distilled water (ddH₂O) and stored at -70 °C. The quality of the DNA and RNA was confirmed by agarose gel electrophoresis and the nucleic acid concentrations were determined spectrophotometrically based on the 260/280 nm absorbance ratio.

Cloning and chromosomal mapping of TaEG

A TaEG EST (DETP-3) was initially identified in the three pistils mutation of wheat by using an annealing con-

trol primer system (ACP) (Yang *et al.*, 2011). Blast search analysis indicated that a full length cDNA (GenBank ID: AK373303) in barley shared perfect identity with the TaEG EST. The PCR primer pair TaEG-1 was designed with Primer Premier 5.0 based on the cDNA of AK373303 (Table 1). The cDNA and genomic DNA sequences were amplified using the same primers (TaEG-1). PCR amplification was done in a thermocycler (My-Cycle, Bio-Rad, San Diego, CA, USA) in a volume of 50 µL containing 100 ng of genomic DNA or reverse transcriptase reaction product (see below for details), 100 µM of each dNTP, 1.5 mM Mg²⁺, 1 U of *Taq* DNA polymerase, 0.4 µM of each primer and 1PCR buffer. The PCR cycling conditions included pre-denaturation at 94 °C for 5 min followed by 35 cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 4 min, and a final extension at 72 °C for 10 min. The amplified products were visualized by gel electrophoresis in 1% agarose gels and then documented with a Gel Doc 2000TM system (Bio-Rad). The target DNA bands were recovered and purified from the gels using Qiaquick Gel extraction kits (QIAGEN, Shanghai, China). The purified PCR products were cloned in the pMD-19T vector according to the manufacturer's instructions (Takata, Dalian, China). Transformants were plated on LB agar containing ampicillin. Clones with inserts were identified using blue/white colony selection. Positive clones were then screened and sequenced by Taihe Biotechnology Co. Ltd. (Beijing, China).

The chromosomal location of TaEG was mapped using the genomic DNA of 36 Dt lines and 5 NT lines (see above). The authenticity of the Dt lines and NT lines was confirmed with SSR markers before use. The gene-specific primers for TaEG-IN3 (Table 1) were designed based on intron which shows clear sequence variations among chromosomes A, B and D (Figure 1).

Molecular characterization of TaEG

The sequence data were analyzed with GenScan software. The open reading frame (ORF) of the cDNA sequence was searched using ORF finder software. A computation tool from the Swiss-Prot/TrEMBL entries was used to calculate the isoelectric point (pI) and molecular mass (Mr) of TaEG. The deduced TaEG sequence was investigated for the presence of signal peptide cleavage sites and transmembrane helices using software available through

Table 1 - Primers used in this study.

PCR target	Primer	Forward	Reverse
Genomic and cDNA sequences	TaEG-1	CCCTTCCGTTCCCAATTCCAAGC	ACATGGCATGGTGATGATACT
Chromosomal location	TaEG-IN3	GGATAATAAGCTCCCTGGTG	TACAATCGTAAGTGCCCTGT
Real-time PCR	TaEG-2	GCGGGCAACGCTGGTCTA	GGGAACATTGGCGGCACA
Internal control	Ubiquitin	AAGGCGAAGATCCAGGACAAG	TGGATGTTGTAGTCCGCCAAG
	Actin	ACGCTTCCTCATGCTATCCTTC	ATGTCTCTGACAATTTCCCGCT

TaEG-4A	GTGTT CAGCTGGGATAATAAGCTCCCTGGTGCACAGGT. AATGTATTAGCGTATCTTGGTTTCTTTAAT	1769
TaEG-4B	GTGTT CAGCTGGGATAATAAGCTCCCTGGTGCACAGGTAAATATATTTGTGTAACCTGGTTTCGTTAAT	1770
TaEG-4D	GTGTT CAGCTGGGATAATAAGCTCCCTGGTGCACAGGT. AATGTATTAGCGTATCTTGGTTTCGTTAAT	1782
TaEG-4A	TCTCAAGACATTCCCATTGCTGACCCATTGGCATTTTAGTTTGCATGATACTCCCTTCTAAACTAATAT	1839
TaEG-4B	TCTTGAGACATTCCCATTGCTGACCCATTGGCATTTTAGTTTGCATGATACTTGTCAAGTAC.	1832
TaEG-4D	TCGCAAGACATTCCCATTGCTGACCCATTGGCATTTTAGTTTGCATGATACTCCCTCCGTAACCTAATAT	1852
TaEG-4A	AGAAGTGTGGATTACTACTTTAGTGATCTAAACGCTCTTATATTAGTTTATAGAGGGAATAC.	1903
TaEG-4B TCCCTCCCTCCCA	1845
TaEG-4D	AAGAGTGTATAGATCACTAAAGTAGTGATCTAAACGCTCTTATATTAGTTTACAGAGTCCCTCCGTCCTCA	1922
TaEG-4A	TAATGTAAGATG. TTTTGGACACTTCTCCCTCCGTTCCGAATTACTTGTCCGAGGTATGGATGTATCTAT	1972
TaEG-4B	TAATGTAAGATGTTT. TTTTGGACACCACAGAGAGTCAA.	1883
TaEG-4D	TAATGTAAGATG. TT.	1936
TaEG-4A	ATGTATTTTAGTTCTAGATACATCCATTTCTGCGACGAGTAATTTGGAACGGAGGGAATACACTAGAGTC	2042
TaEG-4B AAAAACCGTCTTACATTATGGGACCGGAGGGAGTAGATTCCATG	1927
TaEG-4D TTTTGGCACTACACTAGAGTCAAAGCCATCTTACATTATGGGACCGGAGGGAGTAGATTCCATG	2001
TaEG-4A	AAAAACCGTCTGACCGTCAGGTTAGTAGTACTATTGGGACCAAATTTCTTTATAATATCTCACCCAGTTTA	2112
TaEG-4B	TCTAACACACTGACCGTCAGGTTATAGTACTATCGTGACCAAATTTCTTTATAATATGTCACCCAGTTTA	1997
TaEG-4D	TCTAACACACTGACCGTCAGGTTATAGTACTATTGGGACCAAATTTCTTTATAATATCTCACCCAGTTTA	2071
TaEG-4A	CCTGCAGTACTAGTAAAGCGTGTGGTCAACTTGTCTACCATGTGATAAGCATTAAATGCTGACTAAGTAC	2182
TaEG-4B	CCTGCGGTACTAGTAAAGCGTGTAGCCAGCTTGTCTACCATGTGATAAGCATTAAATGCTGCCTAGGTAC	2067
TaEG-4D	CCTGCAGTACTAGTAAAGCGTGTAGTCAACTTGTCTACCATGTGATAAGCATTAAATGCTGCCTAAGTAC	2141
TaEG-4A	AGGGCACTTACGATTGTAGCACAAAAGGTTTCTGCTTGGTGGATATACCAAGGATTCTGTTGAAACCAT	2252
TaEG-4B	AGGGCACTTACGATTGTAGCACAAAAGGTTTCTGCTTGGTGGATATACCTAGGATTCTGTTGAAACCAT	2137
TaEG-4D	AGGGCACTTACGATTGTAGCACAAAAGGTTTCTGCTTGGTGGATATACCAAGGATTCTGTTGAAACCAT	2211

Figure 1 - Alignment of the intron III sequences of TaEG-4A, TaEG-4B and TaEG-4D. The forward and reverse TaEG-IN3 primers used for chromosome location are boxed.

the SignalP 4.1 and SCOP servers. For homology analysis, the amino acid sequence of TaEG was compared with the sequences of other species using Blast 2.1.

Multiple sequence alignment and phylogenetic analysis

The deduced amino acid sequence of TaEG was aligned with the EGase genes reported for other species using the ClustalW program (Thompson *et al.*, 1994). Neighbor-joining trees of the EGase genes were generated with MEGA software version 5.0 (Saitou and Nei, 1987; Tamura *et al.*, 2007) with 1000 bootstrap replicates. Gaps were treated as missing values. The amino acid sequences of EGase genes from *Arabidopsis thaliana* (L.) Heynh (AtCel1: X98544, KOR: NP_199783), *Oryza sativa* L. (OsGLU1: EEE58992), *Solanum lycopersicum* L. (TomCel2: U13055, TomCel3: U78526, TomCel8:

AF098292), *Piper nigrum* L. (CcCel3: X97189), *Nicotiana tabacum* L. (NtCel2: AF362948, NtCel8: AF362949), *Pyrus communis* L. (PeEG2: BAC22691, Cel1: BAF42036), *Mangifera indica* L. (MiCel1: EF608067), *Brassica napus* L. (Cel16: CAB51903) and *Gossypium barbadense* L. (Cel: ADY68791) were obtained from NCBI.

Real-time PCR

Total RNA was isolated from spikes of the near-isogenic line CM28TP and its recurrent parent, CM28, at various developmental stages (2-5, 5-7, 7-10, 10-13, 13-15, 15-17 and 17-20 mm in length). The cDNA was synthesized using a PrimeScript Perfect real-time RT reagent kit (TaKaRa, Dalian, China). The primers for TaEG-2 (Table 1) were designed using Primer Express 2.0 software to amplify 101-bp fragments of the TaEG gene. Real-time as-

says were done with SYBR green dye (TakaRa) using a Bio-Rad CFX96 real-time PCR platform. All of the samples were analyzed in triplicate and the fold change in RNA transcripts was calculated by the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001) with the wheat housekeeping genes ubiquitin (DQ086482) and actin (AB181911) as internal controls (Hama *et al.*, 2004; Yamada *et al.*, 2009).

Results

Cloning and chromosomal location of genomic and cDNA sequences

Cloning followed by PCR product analysis resulted in the identification and isolation of three genomic sequences (3625, 3510 and 3574 bp long) from CM28TP. Three bands (489, 435 and 374 bp) were amplified in CM28TP with gene-specific primers of TaEG-IN3. NT and Dt line analysis indicated that the TaEG-IN3 primer set amplified the expected three bands from all the NT and Dt lines, with the exception of DT4AL (lacked the 489-bp band), N4A-T4B (lacked the 489-bp band), Dt4BS (lacked the 374-bp band) and Dt4DS (lacked the 435 bp band) (Figure 2). Further analysis showed that the longest PCR product (3625 bp) was located on the long arm of chromosome 4A, the intermediate product (3574 bp) on the short arm of chromosome 4D and the shortest product (3510 bp) on the short arm of chromosome 4B. The three homologous genes were tentatively designated as TaEG-4A, TaEG-4D and TaEG-4B,

respectively. The sequences were deposited in GenBank under accession numbers KC521526, KC521527 and KC521528.

The cDNA from young spikes of CM28TP amplified with the TaEG-1 primer pair yielded a fragment of ~2.1 kb. The PCR product was cloned into the PMD-19T vector and 15 clones were sequenced. Three distinct sequences of 2061 bp, 2084 bp and 2087 bp were identified (accession numbers: AC521523, AC521524 and AC521525, respectively), indicating that the band of ~2.1 kb included co-migrating cDNA of different homologous genes. The three homologous genes shared 100% homology in their coding and 3' untranslated regions. The lengths of the 5' untranslated region were 103 bp, 126 bp and 129 bp for TaEG-4A, TaEG-4B and TaEG-4D, respectively. The open reading frame (ORF) of TaEG was 1869 bp and coded for a deduced protein of 622 amino acids. Primary structure analysis using Swiss-Prot/TrEMBL revealed that the molecular mass of the putative TaEG protein was 69 kDa with a theoretical pI = 9.39. The putative TaEG protein lacked signal peptide cleavage sites but had a 74~96 amino acid transmembrane domain. BLAST searches showed that the TaEG belonged to glycosyl hydrolase family 9 (Henrissat *et al.*, 1989). Sequence alignment suggested that TaEG had high homology with other plant EGase proteins, all of which contained four Cys residues and two putative glycosyl hydrolase active sites, namely, 'SY-VG-G-YP-VHHR' and 'F-DVR-N-NYTE-TLAGAN' (Figure 3).

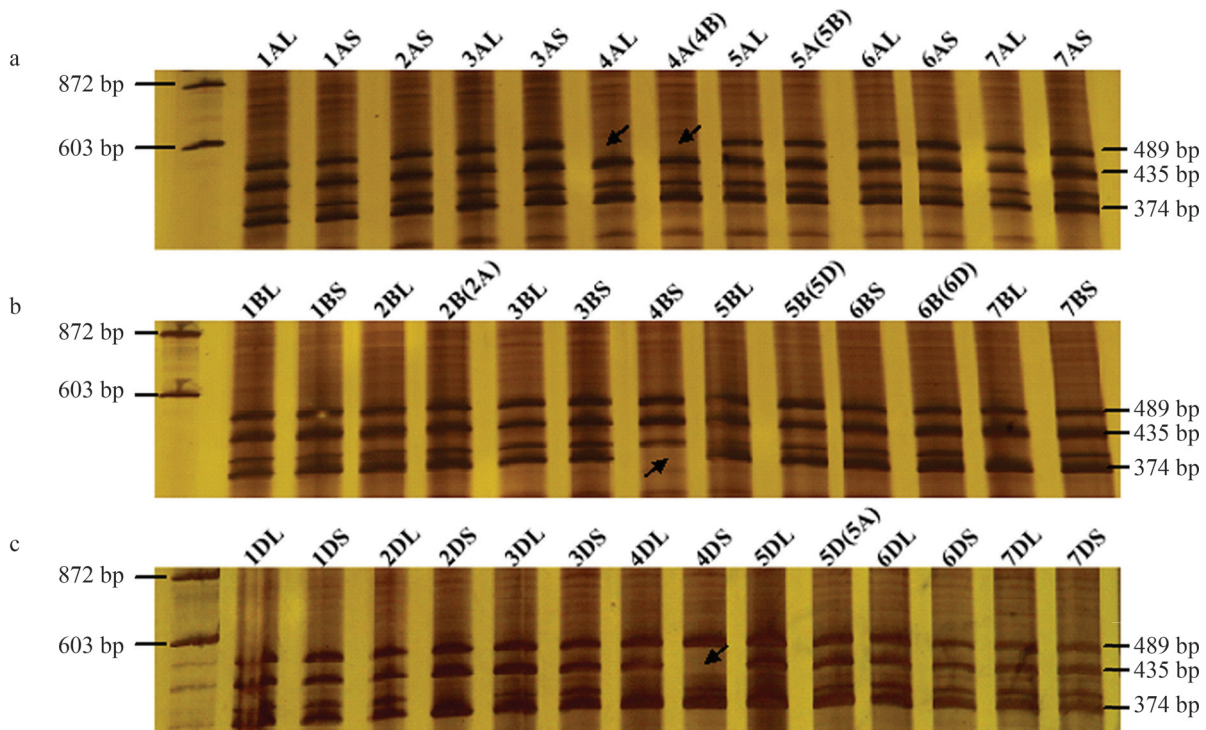


Figure 2 - Chromosome mapping of the TaEG gene. DNA size markers are shown on the left of the gel. (A) TaEG-4A located on the long arm of chromosome 4A. (B) TaEG-4B located on the short arm of chromosome 4B. (C) TaEG-4D located on the short arm of chromosome 4D. Missing fragments are indicated by arrows.

CEL1	MSMYGRDPWGGPLEINTTDSATDDDRSRNLQDLRAALSSRPLDETQQSWLLGPGS. EQ. KKKKYVDLGC I VSRK I FVWTVGTLLVSAFLAGFI	93
CEL16	.. MYGRDPWGGPLE I HATDSATDDDRSRNLNDI DRAAL. SRPLDETQQSWLLGPT. QK. KKKKYVDLGC I VSRK I FVWTVGT I VAA. ALLAGFI	89
CEL	.. MYGRDPWGGPLE I NATDSATDDDRSRNLQDLRAAL. SRPLDETQQSWLLGPG. QK. KKKKYVDLGC I VSRK I FVWTVGTLLVSAFLAGFI	90
OSGLU1	MSMYGRDPWGGPLE I CH. DSATDDDRSRNL. DLDRGAL. SRTLDETQQSWLLAGPGDQGRKKKYVDLGC I VSRK I FVWTVGTLLVAAVAVAGLV	92
KOR	.. MYGRDPWGGPLE I NTADSATDDDRSRNLNDL DRAAL. SRPLDETQQSWLLGPT. QK. KKKKYVDLGC I VSRK I FVWTVGT I VAA. ALLAGFI	89
TOMCEL3	.. MYGRDPWGGPLE I HTADSATDDDRSRNLQDFDRAAM. SRSLDETQQSWLLGPT. Q. . KKKKYVDLGC I VSRK I FKWTGVC I I AALLAGFI	89
TAEG	MSMYGRDPWGGPLE I CH. DSATDDDRSRNL. DL DRAAL. SRTLDETQQSWLLAGPGDQGRKKRYVDLGC I VSRK I FVWTVGTLLVAAVAVAGLV	92
CEL1	TL I VKT VPRHRHRSPPDNYTLALRKSLMFFNAQRSGKLPKHNNVSWRGNSSLRDN. DPSTTFKDLAGGFYDAGDA I KFNFPASFAMTMLSWSV	187
CEL16	TL I VKT VPRHRHRTKPPDNYT I ALHKALKFFNAQKSGKLPKHNNVSWRGNSSLQDGKDGSGSFYKDLVGGYDAGDA I KFNFPAMAYMTMLSWSV	184
CEL	TL I VKT VPRHHHRHRSPPDNYTLALHKALMFFNAQRSGKLPKHNNVSWRGNSSLQDGKSDPSVLMKDLVGGYDAGDA I KFNFPASFAMTMLSWSV	185
OSGLU1	AG I AKA I PRHRPPPPDDFTVALRKALMFFNAQKSGKLPKHNNVSWRGNSSCMKDG L S DPAVGRS. LVGGYDAGDAVKFNFPAAFAMTMLSWSV	186
KOR	TL I VKT VPRHHPKTPPDNYT I ALHKALKFFNAQKSGKLPKHNNVSWRGNSSLQDGKGETGSFYKDLVGGYDAGDA I KFNFPAMAYMTMLSWSV	184
TOMCEL3	TM I VKLVPRKHHPDNYTLALRKALMFFNAQKSGKLPKHNNVSWRGNSSLQDGKSDSTMFKNLVGGYDAGDA I KFNFPQSFALMTMLSWSV	184
TAEG	AG I AKA VPRHQAPPAPLDQFTLALRKALMFFNAQKSGKLPKHNNVSWRGNSSCLDKGRSDQTSRRD. LSGGYDGGSAVKFNFPASFAMTMLSWSV	186
CEL1	IEYSAKYEAAGELSHVKE I I KWGSDYFLKTFNHSADS I DRLVAQVGVSTAGGSTPNPDHYCWMRPED I DYQRPVYQCHS. CSDLAAEMAALAA	281
CEL16	IEYSAKYEAAGELVHVKE I I KWGTDYFLKTFNSTADS I DDLVSVQVSGNTDDGSDPNPDHYCWMRPEDMDYKRPVTTGNGG CSDLAAEMAALAS	279
CEL	IEYSAKYEAAGELNHVKE I I KWGTDYLLKTFNNTADT I DRIAAQVGI DGTSGGSSAPNDHYCWMRPED I DYPRPVVECHS. CSDLAAEMAALAS	279
OSGLU1	IEYSAKYEAAGELGHI RDT I KWGADYFLKTFNSTADT I DRVVMQVSGATSPGSTQPNPDHYCWMRPED I DYPRPVVECHA. CSDLAAEMAALAA	280
KOR	IEYSAKYEAAGELTHVKE I I KWGTDYFLKTFNSTADS I DDLVSVQVSGNTDDGSDPNPDHYCWMRPEDMDYKRPVTTGNGG CSDLAAEMAALAS	279
TOMCEL3	IEYSAKYEAAGELAHVKDT I KWGTDYLLKTFNNSADT I DRIAAQVGGDTTGGATPNPDHYCWMRPED I DYARPVTECHG. CSDLAAEMAALAS	278
TAEG	IEYNAKYEAAGELGHRDT I KWGADYLLKTFNSTARS I DRLVAQVGGAAAMSEGPSQPNPDHYCWMRPED I DYPRPVTECHT. CSDLAAEMAALAA	280
CEL1	SS I VFKDNKAYSQKLVHGAKTLFRFSREQ. . RGRYSAGGSTDAATFYNSTSYWDEF I WGGAWMYATGNSSYLQLATTPGLAKHAGAFWGGPDYG	374
CEL16	AS I VFKDNREYSKLVHGAKTVYQFGRTR. . RGRYSAGTA. ESAKFYNSSMYWDEF I WGGAWLYYATGNVTYLDL I TKPTMAKHAGAFWGGPYG	371
CEL	AS I VFKDNKAYSQKLVHGARTLKFARQ. . RGRYSAGGS. DPALFYNSSSYWDEFWGGAWLYYATGNSSYLQLATHPKLAKHAGAFWGGPDYG	371
OSGLU1	AS I VFKDNKAYSQKLVHGATTLKFARQN. . RGRYSAGGS. DAAKFYNSTSYWDEFWGGSWMYLATGNSSYLQLATHPKLAKHAGAFWGGPDYG	372
KOR	AS I VFKDNKEYSKLVHGAKVYQFGRTR. . RGRYSAGTA. ESSKFYNSSMYWDEF I WGGAWMYATGNVTYLNL I TQPTMAKHAGAFWGGPYG	371
TOMCEL3	AS I VFKDNKAYSQKLVHGARTLKFASRDQ. . RGRYSV. GN. EAETFYNSTGYWDEF I WGAAWLYYATGNSSYLQLATTPGI AKHAGAFWGGPDYG	369
TAEG	AS I VFRDNRAYSHKLLHGATTVWDFARKGGRQTSVPRS. DAAKFYNSTAYWDEY I WGGSWMYLATGNSSYLQFATDTKLAKNAH I YSRPPNYG	374
CEL1	VLSWDNKLGAQVLLSRLRFLSPGYPIE I LRTFHNGTS I I M C S Y L P V F T F N R T K G G L I Q L N H G R P Q P L Q Y V V N A A F L A T L Y S E Y L D A A D T P G	469
CEL16	VFSWDNKLGAQVLLSRLRFLSPGYPIE I VRTFHNGTS I I M C S Y L P F N K F N R T R G G L I E L N H G D P Q P L Q Y A A N A A F L A T L Y S D Y L D A A D T P G	466
CEL	VLSWDNKLGAQVLLSRLRFLSPGYPIE I LSTFHNGTS I I M C S F L P V F T S F N R T K G G L I Q L N H G R P Q P L Q Y V V N A A F L A A L Y S D Y L D A A D T P G	466
OSGLU1	VFSWDNKLGAQVLLSRLRFLSPGYPIE I LRTFHNGTS I I M C S Y L P I F K S F N R T K G G L I Q L N H G R P Q P L Q Y V V N A A F L A S L Y G D Y L E A A D T P G	467
KOR	VFSWDNKLGAQVLLSRLRFLSPGYPIE I LRTFHNGTS I I M C S Y L P I F N K F N R T N G G L I E L N H G A P Q P L Q Y S V N A A F L A T L Y S D Y L D A A D T P G	466
TOMCEL3	VLSWDNKLGAQVLLSRLRFLSPGYPIE I LRTFHNGTS I I M C S Y L P I F T S F N R T K G G L I Q L N H G R P Q P L Q Y V V N A A F L A T L F S D Y L A A A D T P G	464
TAEG	VFSWDNKLGAQVLLSRLRFLSPGYPIE I LRTYHNQTS I I M C S Y L P N F R S F N R T K G G L I Q L N H G Q P Q P L Q Y V V N A A F P A S L F S D Y L E A A D T P G	469
CEL1	WYCGPNFYSTDVLRFAETQ I D Y I L G K N P R K M S Y V V G F G N H Y P K H V H R G A S I P K N . K I K Y N C K G G W K W R D T P K A N P N T I G A M V A G P D K H D G F R	563
CEL16	WYCGPNFYSTNVLRFAETQ I D Y I L G K N P R K M S Y L V G F G T K Y P K H V H R G A S I P K N . K V K Y N C K G G W K W R D S K P N P N T I E G A M V A G P D K R D G Y R	560
CEL	WYCGPNFYSTDVLRFAETQ I D Y I L G K N P R K L S Y V V G F G N H Y P K H V H R G A S I P K N . K I K Y N C K G G W K W R D T S K P N P N T L V G A M V A G P D K H D G F R	560
OSGLU1	WYCGPHFY I E T L R N F A R T Q I E Y I L G K N P L K M S Y V V G Y G N R Y P K R V H H R G A S I P K N . G V H Y C K K G G W K W R E T K P N P N I I V G A M V A G P D R H D G F K	561
KOR	WYCGPNFYSTVLRDFARSQ I D Y I L G K N P R K M S Y V V G F G T K Y P R H V H R G A S I P K N . K V K Y N C K G G W K W R D S K P N P N T I E G A M V A G P D K R D G Y R	560
TOMCEL3	WYCGPNFYSTDVLRKFAETQ I D Y I L G K N P R K M S Y V V G F G N H Y P K H V H R G A S I P K N . K V K Y N C K G G W K Y R D S S K A N P N T I V G A M V A G P D K H D G F R	558
TAEG	WYCGPHFY I E V L R S F A R T Q M E Y I L G K N P L K M S Y V V G H G K H Y P K R V H H R G A S I P K K K G V H P C K G G W T W R D T K P N P N I I V G A M V A G P D R H D G F K	564
CEL1	DVRSNRYNYTEPTLGNAGLVAALVALSGEKS. . I D K N T I F S A V P P M F P T P P P P A P W K P	622
CEL16	DVRTNRYNYTEPTLGNAGLVAALVALSGEEASG I D K N T I F S A V P P L F P T P P P P A P W K P	621
CEL	DVRTNRYNYTEPTLGNAGLVAALVALSGDKATG. . I D K N T I F S A V P P M F P T P P P P A P W K P	619
OSGLU1	DVRKNRYNYTEATLGNAGLVAALVALSGEGH. G. . V D K N T M F S A V P P M F P S P P P P A P W K P	619
KOR	DVRMRYNYTEPTLGNAGLVAALVALSGEEATG I D K N T I F S A V P P L F P T P P P P A P W K P	621
TOMCEL3	DVRSNRYNYTEPTLGNAGLVAALVALSGDRDVG. . I D K N T L F S A V P P M F P T P P P P A P W K P	617
TAEG	DVRRNRYNYTEPTLGNAGLVAALVALSGEGH. G. . V D K N T M F S A V P P M F P S P P P P A P W K P	622

Figure 3 - Alignment of the deduced amino acid sequence of TaEG with other plant EGase sequences. Four Cys residues and two putative glycosyl hydrolase active sites are indicated by the dotted line and solid line frames, respectively.

Structure of the TaEG genomic sequences

The TaEG-4A, TaEG-4B and TaEG-4D genomic sequences were 3625 bp, 3510 bp and 3574 bp long, respectively; their alignment and comparison with the corresponding cDNA sequences revealed the complex structure of the three homologous genes that consisted of six exons and five introns (Figure 4). The five introns fulfilled the GT-AG rule but varied markedly in length and sequence composition. The length of intron III was 905 bp, 792 bp and 850 bp in TaEG-4A, TaEG-4B and TaEG-4D, respectively, which was longer than the other four introns. Intron IV was 85 bp long in the three genes and was the shortest of the five introns. Intron I was 395 bp, 369 bp and 376 bp long in TaEG-4A, TaEG-4B and TaEG-4D, respectively. Intron II showed two lengths (88 bp and 86 bp) in TaEG-4A, TaEG-4B and TaEG-4D whereas intron V had three lengths, *i.e.*, 91 bp, 92 bp and 90 bp.

Phylogenetic analysis of TaEG and plant EGases

To construct the phylogenetic tree, EGases from *A. thaliana* (AtCel1 and KOR), *O. sativa* (OsGLU1), *S. lycopersicum* (TomCel2, TomCel3 and TomCel8), *P. nigrum* (CcCel3), *N. tabacum* (NtCel2 and NtCel8), *P. communis* (PcEG2 and Cel1), *M. indica* (MiCel1), *B. napus* (Cel16) and *G. barbadense* (Cel) were analyzed. Figure 5 shows the tree obtained with the neighbor joining program using MEGA software and 1000 bootstrap replicates. Previous analysis identified two types of EGase in plants, namely, secretory EGase and transmembrane EGase, with the major difference between them being the presence of a signal peptide in the former and a transmembrane domain in the latter (Brummell *et al.*, 1997a). In the present study, these EGases formed into two well-resolved clades (clades 1 and 2). Clade 1 contained eight members and the deduced proteins possessed a signal peptide, *e.g.*, Tom Cel2, NtCel2 and PcEG2. Further inspection of clade 1 revealed the presence of two subclades (subclades 1 and 2), with the major difference between them being the presence (subclade 1) or absence (subclade 2) of a cellulose binding domain (CBD) (Zhou *et al.*, 2004). Clade 2 contained seven members and the deduced proteins had a transmembrane domain, *e.g.*, Tom OsGLU1, Cel16 and KOR. TaEG was located in clade 2 and showed a much greater similarity (83%) to *O. sativa*

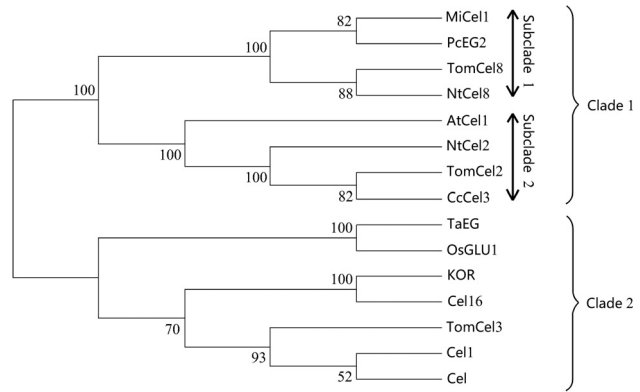


Figure 5 - Phylogenetic tree of TaEG and other homologous sequences. The numbers at the nodes indicate bootstrap values.

OsGLU1. Phylogenetic analysis also indicated that TaEG was a membrane-anchored protein.

TaEG gene expression in developing spikes

Real-time PCR was used to study the pattern of TaEG expression in different tissues during the developmental stages of young spikes. As shown in Figure 6, in young spikes of CM28TP at stages 2-5, 5-7, 10-13, 13-15, 15-17 and 17-20 mm the transcript levels were similar to those observed in CM28. However, the transcript levels in young spikes of CM28TP at stages 7-10 mm were 5.6-fold higher than in CM28 spikes at the same stage.

Discussion

An EST sequence (DETP-3) previously identified in the three pistil mutation of wheat using an annealing control primer system (ACP) was shown to have functions similar to those of the EGase gene (Yang *et al.*, 2011). The EST was therefore tentatively designated as an *T. aestivum* endoglucanase gene (TaEG). In the present study, cDNA cloning and genome sequencing of TaEG from wheat showed that there are three homoeologous TaEG genes in wheat, namely, TaEG-4A, TaEG-4B and TaEG-4D located on chromosomes 4AL, 4BS and 4DS, respectively. The three homoeologous genes showed high conservation of the nucleotide coding sequences and 3 untranslated regions, with homologies up to 100%. The putative TaEG

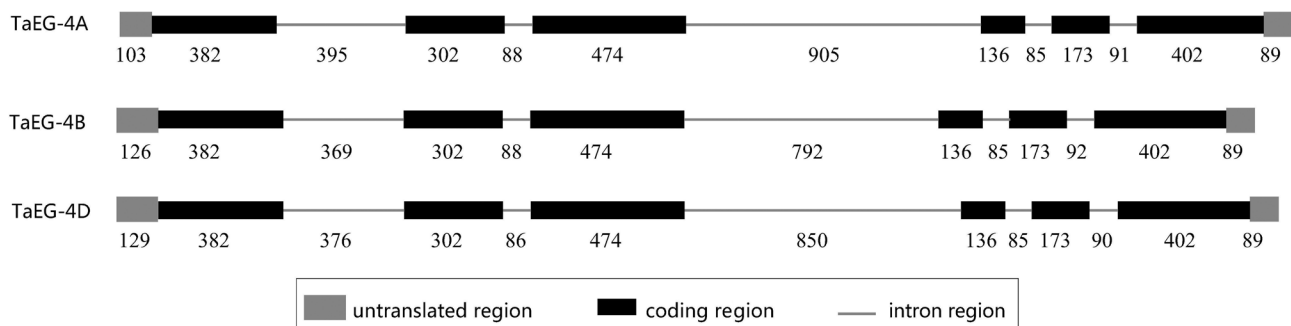


Figure 4 - Schematic representation of the TaEG gene.

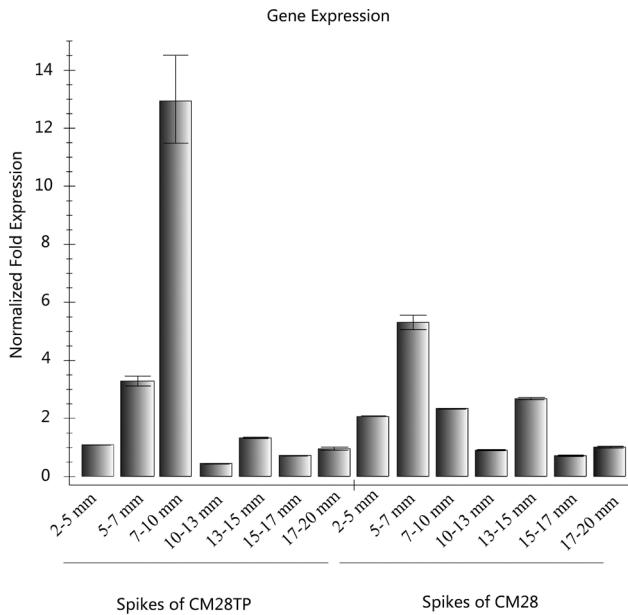


Figure 6 - TaEG expression assessed by real-time PCR. Total RNA was isolated from spikes of the near-isogenic line CM28TP and its recurrent parent CM28 at various developmental stages. Ubiquitin and actin were used as internal controls. The transcript levels are shown as relative values and the columns represent the mean \pm SEM of three replicates.

protein had a Mr of 69 kDa, a theoretical pI of 9.39 and a 74-96 amino acid transmembrane domain in the N-terminus. Despite variation in the nucleotide and amino acid sequences of TaEG from different plants, all of these proteins contained four Cys residues and two putative glycosyl hydrolase active sites (SY-VG-G-YP-VHHR and F-DVR-N-NYTE-TLAGAN) (Figure 3). The genome sequences of TaEG-4A, TaEG-4B and TaEG-4D contained six exons and five introns, all of which (except for intron IV) varied in length and sequence composition.

Endo- β -1,4-glucanases have been classified into two subfamilies (Nicol *et al.*, 1998). One subfamily consists of soluble secreted enzymes that contain an N-terminal signal peptide and EGase domain; these enzymes are located in the periplasm where they function as cell wall-softening enzymes that modify the cell wall. The other EGase subfamily consists of membrane-anchored proteins that have a predicted N-terminal transmembrane anchor motif (Brummell *et al.*, 1997a). TaEG, like KOR and TomCel3, belongs to the latter subfamily and contains a transmembrane domain located between amino acids 74 to 96. The precise function of membrane-anchored EGases is currently unclear, but these enzymes are located in the plasma membrane and participate in cellulose metabolism in the inner layers of the cell wall (Brummell *et al.*, 1997a; Nicol *et al.*, 1998; Del Campillo, 1999; Molhoj *et al.*, 2001). Some membrane-anchored EGases are involved in cell elongation. In tomato vegetative tissues the highest levels of TomCel3 mRNA were found in the most actively growing cells, in the expanding zones of young hypocotyls and in

young expanding leaves (Brummell *et al.*, 1997b). The *Arabidopsis* KOR gene has been identified with the dwarf mutant KORRIGAN and found to be involved in normal wall assembly and cell elongation (Nicol *et al.*, 1998). In *G. barbadense* the Cel gene is involved in cotton fiber elongation (Zhu *et al.*, 2011). Our phylogenetic analysis revealed that the TaEG genes clustered more closely with *O. sativa* OsGLU1. Furthermore, alignment analysis demonstrated that TaEG shared more identity with OsGLU1 (83%) than other membrane-anchored EGases. These findings suggest that TaEG is an ortholog of *O. sativa* OsGLU1 and may have similar biological roles. The *O. sativa* OsGLU1 gene has been identified with the dwarf mutant *glu* and found to be involved in cell elongation, cellulose and pectin content (Zhou *et al.*, 2006).

Expression analysis showed that the highest levels of TaEG mRNA occurred in the 7-10 mm stage of CM28TP spikes. CM28TP has two more pistils per floret than CM28 and young spikes at the 7-10 mm stage in the primordial differentiation phase of subsidiary pistils (Yang *et al.*, 2011). Based on these observations, we suggest that TaEG may contribute to pistil development and could be functionally important in defining the morphology of the three pistil mutant. However, this suggestion requires confirmation through additional experiments.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant no. 30871533) and the Scientific Research Fund of Sichuan Provincial Education Department (grant no. 12ZB325).

References

- Beguín P (1990) Molecular biology of cellulose degradation. *Annu Rev Microbiol* 44:219-248.
- Beguín P and Aubert JP (1994) The biological degradation of cellulose. *FEMS Microbiol Rev* 13:25-58.
- Bisaria VS and Mishra S (1989) Regulatory aspects of cellulase biosynthesis and secretion. *Crit Rev Biotechnol* 9:61-103.
- Brummell DA, Bird CR, Schuch W and Bennett AB (1997b) An endo-1,4- β -glucanase expressed at high levels in rapidly expanding tissues. *Plant Mol Biol* 33:87-95.
- Brummell DA, Catala C, Lashbrook CC and Bennett AB (1997a) A membrane-anchored E-type endo-1,4- β -glucanase is localized on Golgi and plasma membranes of higher plants. *Proc Natl Acad Sci USA* 94:4794-4799.
- Byrne KA, Lehnert SA, Johnson SE and Moore SS (1999) Isolation of a cDNA encoding a putative cellulase in the red claw crayfish *Cherax quadricarinatus*. *Gene* 239:317-324.
- Christoffersen RE, Tucker ML and Laties GG (1984) Cellulase gene expression in ripening avocado fruit: The accumulation of cellulase mRNA and proteins as demonstrated by cDNA hybridization and immunodetection. *Plant Mol Biol* 3:385-391.
- Del Campillo E (1999) Multiple endo-1,4- β -D-glucanase (cellulase) genes in *Arabidopsis*. *Curr Top Dev Biol* 46:39-61.

- Endo TR and Gill BS (1996) The deletion stocks of common wheat. *J Hered* 87:295-307.
- Ferrarese L, Trainotti L, Moretto P, Polverino de Laureto P, Rascio N and Casadoro G (1995) Differential ethylene-inducible expression of cellulase in pepper plants. *Plant Mol Biol* 29:735-747.
- Hama E, Takumi S, Ogihara Y and Murai K (2004) Pistillody is caused by alterations to the class-B MADS-box gene expression pattern in alloplasmic wheats. *Planta* 218:712-720.
- Henrissat B, Claeysens M, Tomme P, Lemesle L and Mornon JP (1989) Cellulase families revealed by hydrophobic cluster analysis. *Gene* 81:83-95.
- Kemmerer EC and Tucker ML (1994) Comparative study of cellulases associated with adventitious root initiation, apical buds, and leaf, flower and pod abscission zones in soybean. *Plant Physiol* 104:557-562.
- Lane DR, Wiedemeier A, Peng L, Hofte H, Vernhettes S, Desprez T, Hocart CH, Birch RJ, Baskin TI, Burn JE, *et al.* (2001) Temperature-sensitive alleles of RSW2 link the KORRIGAN endo-1,4- β -glucanase to cellulose synthesis and cytokinesis in *Arabidopsis*. *Plant Physiol* 126:278-288.
- Lashbrook CC, Gonzalez-Bosch C and Bennett AB (1994) Two divergent endo- β -1,4-glucanase genes exhibit overlapping expression in ripening fruit and abscising flowers. *Plant Cell* 6:1485-1493.
- Livak KJ and Schmittgen TD (2001) Analysis of relative gene expression data using real time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* 25:402-408.
- Manickavelu A, Kambara K, Mishina K and Koba T (2007) An efficient method for purifying high quality RNA from wheat pistils. *Colloids Surf B Biointerfaces* 54:254-258.
- Molhoj M, Jorgensen B, Ulvskov P and Borkhardt B (2001) Two *Arabidopsis thaliana* genes, KOR2 and KOR3, which encode membrane-anchored endo-1,4- β -D-glucanases, are differentially expressed in developing leaf trichomes and their support cells. *Plant Mol Biol* 46:263-275.
- Nicol F, His I, Jauneau A, Vernhettes S, Canut H and Hofte H (1998) A plasma membrane-bound putative endo-1,4- β -D-glucanase is required for normal wall assembly and cell elongation in *Arabidopsis*. *EMBO J* 17:5563-5576.
- Peng ZS (2003) A new mutation in wheat producing three pistils in a floret. *J Agro Crop Sci* 189:270-272.
- Porebski S, Bailey LG and Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol Biol Rep* 1:8-15.
- Rosso MN, Favery B, Pottie C, Arthaud L, De Boer JM, Hussey RS, Bakker J, Baum TJ and Abad P (1999) Isolation of a cDNA encoding a β -1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. *Mol Plant Microbe Interact* 12:585-591.
- Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
- Sears ER (1954) The aneuploids of common wheat. *Mo Agric Exp Stn Res Bull* 572:1-58.
- Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software ver. 4.0. *Mol Biol Evol* 24:1596-1599.
- Taylor JE, Coupe SA, Picton S and Roberts JA (1994) Characterization and accumulation pattern of an mRNA encoding an abscission-related β -1,4-glucanase from leaflets of *Sambucus nigra*. *Plant Mol Biol* 24:961-964.
- Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680.
- Tucker ML, Sexton R, Del Campillo E and Lewis LN (1988) Bean abscission cellulase: Characterization of a cDNA clone and regulation of gene expression by ethylene and auxin. *Plant Physiol* 88:1257-1262.
- Watanabe H, Noda H, Tokuda G and Lo N (1998) A cellulase gene of termite origin. *Nature* 394:330-331.
- Yamada K, Saraike T, Shitsukawa N, Hirabayashi C, Takumi S and Murai K (2009) Class D and B sister MADS-box genes are associated with ectopic ovule formation in the pistil-like stamens of alloplasmic wheat (*Triticum aestivum* L.). *Plant Mol Biol* 71:1-14.
- Yang ZJ, Peng ZS, Wei SH, Yu Y and Cai P (2011) Identification of differentially expressed genes in three pistils mutation in wheat using annealing control primer system. *Gene* 485:81-84.
- Yang ZJ, Peng ZS, Zhou YH, Peng LJ and Wei SH (2012) Evaluation on the genetic background of wheat near isogenic lines for three pistils character by SRAP markers. *J Nucl Agri Sci* 1:22-27.
- Zhou W, Takeda H, Liu XZ, Nakagawa N, Sakurai N, Huang J and Li YQ (2004) A novel endo-1,4- β -glucanase gene (*LlpCell1*) is exclusively expressed in pollen and pollen tubes of *Lilium longiflorum*. *Acta Bot Sin* 46:142-147.
- Zhou HL, He SJ, Cao YR, Chen T, Du BX, Chu CC, Zhang JS and Chen SY (2006) OsGLU1, a putative membrane-bound endo-1,4- β -D-glucanase from rice, affects plant internode elongation. *Plant Mol Biol* 60:137-151.
- Zhu HY, Han XY, Lv JH, Zhao L, Xu XY, Zhang TZ and Guo WZ (2011) Structure, expression differentiation and evolution of duplicated fiber developmental genes in *Gossypium barbadense* and *G. hirsutum*. *BMC Plant Biol* 11:40-54.

Internet Resources

- GenScan software, <http://genes.mit.edu/GENSCAN.html> (accessed November 7, 2012).
- ORF finder software, <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>.
- Swiss-Prot/TrEMBL tool for calculating isoelectric point and molecular mass, http://www.expasy.org/tools/pi_tool.html (accessed November 7, 2012).
- SignalP 4.1 software for identifying signal peptide cleavage sites, <http://www.cbs.dtu.dk/services/SignalP/> (accessed November 7, 2012).
- SCOP software for detecting transmembrane helices, <http://scop.mrc-lmb.cam.ac.uk/scop/> (accessed November 7, 2012).

Associate Editor: Houtan Noushmehr