DOI: 10.1007/s10535-018-0775-6 BIOLOGIA PLANTARUM 62 (2): 369-378, 2018

Characterization of novel D-hordeins from *Psathyrostachys juncea*

X.K. HU[1,2], S.F. DAI[1], T. OUELLET[2], M. BALCERZAK[2], H. ROCHELEAU[2], S. KHANIZADEH[2], Z.J. PU[3], and Z.H. YAN[1]*

Triticeae Research Institute, Sichuan Agricultural University, Chengdu 611130, Sichuan, P.R. China[1]
Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON K1A 0C6, Canada[2]
Crop Research Institute, Sichuan Academy of Agricultural Science, Chengdu 610066, Sichuan, PR China[3]

Abstract

Three genes encoding novel D-hordeins, *Ns 1.3*, *Ns 2.6*, and *Ns 2.9* were isolated from *Psathyrostachys juncea*. The Ns 1.3 differed from Ns 2.6 and Ns 2.9 by having a shorter open reading frame (< 1.5 kb *versus* > 2.5 kb) and was probably not expressed as a normal protein, while the activities for Ns 2.6 and Ns 2.9 were verified by bacterial expression. Though highly similar primary structure to wheat high molecular mass glutenin subunits (HMM-GSs) and barley D-hordeins, Ns 2.6 and Ns 2.9 had more cysteine residues (nine in total) and a larger molecular mass than HMM-GSs, and a longer N-terminal length than D-hordeins. Phylogenetic analysis revealed that the *Ps. juncea* D-hordeins were divided into Ns 1.3 type and Ns 2.6/Ns 2.9 type. Divergence times indicated that *Ns 1.3* diverged the earliest from the orthologous *Triticeae* locus, while *Ns 2.6* and *Ns 2.9* and the D-hordeins from two *Hordeum* species diverged nearly at the same time from those loci, and the divergence between the D-hordeins of *H. chilense* and *Ns 2.6/Ns 2.9* was more recent than between the two *Hordeum* species. The novel *Ps. juncea* D-hordeins have the potential to be very important for improving the end-use quality of wheat flour because of the presence of extra cysteine residues and longer repetitive domain, in addition they can contribute to the understanding of the evolution of *Triticeae* prolamins.

Additional key words: bacterial expression, *Hordeum chilense*, *H. vulgare*, HMM-prolamins, phylogenetic analysis, *Triticum aestivum*.

Introduction

Prolamins are important storage proteins in the endosperms of barley, wheat, and related species (Shewry and Halford 2002). Based on the amino acid sequence, prolamins can be further classified into three different groups: sulphur-rich, sulphur-poor, and high molecular mass (HMM) prolamins (Shewry and Tatham 1990). The wheat HMM-prolamins are high molecular mass glutenins (HMM-GS), while barley HMM-prolamins are D-hordeins (Gu *et al*. 2003, Pistón *et al*. 2007).

HMM-GSs are key components of storage proteins in wheat and closely related wild grasses (Shewry *et al*. 2003). Hexaploid wheat has three sets of paralogous genes that control the biosynthesis of HMM glutenin proteins. These genes are mapped to the long arms of group one chromosomes at three homoeologous loci (*Glu-A1*, *Glu-B1*, and *Glu-D1*) in the wheat A, B, and D genomes. Each locus expresses a large x- and small y-type subunits that originate from two closely related x- and y-type genes (Payne 1987). HMM glutenins enhance the quality of wheat flour and account for ~10 % of total endosperm proteins (Rasheed *et al*. 2014). In order to expand the gene pools that could improve wheat flour quality as well as to better understand the subtle variations in gene origin and evolution, most researches have focused on HMM-prolamin genes from wild relatives of wheat (Wan *et al*. 2002, Liu *et al*. 2010, Dai *et al*. 2013, Zhang *et al*. 2016).

Despite recent progress in isolating HMM-prolamin

Submitted 15 March 2017, *last revision* 19 August 2017, *accepted* 24 August 2017.
Abbreviations: AA - amino acid; CTAB - cetyltrimethyl ammonium bromide; DTT - dithiothreitol; HMM-GSs - high molecular mass glutenin subunits; IPTG - isopropyl-β-D-1-galactoside; MYA - million years ago; ORF - open reading frame; PCR - polymerase chain reaction; SDS-PAGE - sodium dodecyl sulfate polyacrylamide gel electrophoresis; UTR - untranslated region.
Acknowledgements: This work was supported by the Ministry of Science and Technology of China (2016YFD0100500), the National Natural Science Foundation of China (U1403185), the Key Fund Project of the Sichuan Provincial Department of Education (15ZA0021), and the Agriculture and Agri-Food Canada's Genomics Research and Development Initiative; Pu ZJ was supported by the Sichuan Province Youth Science and Technology Innovation Team (2014TD0014), Hu XK was supported by the China Scholarship Council under the MOE-AAFC PhD Research Program .
* Corresponding author; fax: (+86) 2882650350, e-mail: zhyan104@163.com

genes from wheat species, there is little knowledge regarding orthologous HMM-prolamins from distantly related cereal species, such as *Hordeum vulgare*, *H. chilense,* and *Psathyrostachys juncea* and *Ps. fragilis*. Previous studies have indicated that barley D-hordeins are structurally related to wheat HMM glutenins and are encoded by the *Hor3* locus on chromosome 5 (Blake *et al*. 1982, Halford *et al*. 1992). The diversity of D-hordein alleles has been investigated from barley collections. Atanassov *et al*. (2001) analyzed the D-hordeins of 49 naked barley accessions, revealing 4 unique alleles, whereas Yin *et al*. (2003) investigated the variation of D-hordeins from 211 accessions of close undomesticated relatives of barley (*H. vulgare* ssp. *agriocrithon* and ssp. *spontaneum*), and 13 different alleles were identified. D-hordeins are very small proteins, and their electrophoretic mobility is intermediate between the high molecular mass glutenin subunits (HMM-GSs) Bx7 and Dy12 of hexaploid wheat (Yin *et al*. 2003). These D-hordeins variations are interesting traits that could be transferred into *Tritordeum*, a novel cereal that was bred by distant hybridization between tetraploid/hexaploid wheat and barley species, which contributes to different bread-making and malting quality, as D-hordeins are known to be closely related to wheat HMM-GSs (Alvarez *et al*. 1992, Martín *et al*. 1999). The primary sequences of D-hordeins and HMM-GSs are very similar, although the former contains more cysteine residues and longer repetitive regions than the latter. Thus, D-hordeins have been restricted to *Hordeum* species and the closely related, artificially-developed new cereal *Tritordeum* (Alvarez *et al*. 1999, Gu *et al*. 2003, Pistón *et al*. 2007).

Psathyrostachys is a perennial *Triticeae* genus that is comprised of eight species (Baden 1991). Species within the *Psathyrostachys* genus share the Ns genome, which is distantly related to the A, B, and D genomes of wheat, but does not have any genomic relationship with the *Triticeae* I, H, R, St, P, E, and W genomes (Hsiao *et al*. 1986, Baden *et al*. 1989). *Psathyrostachys* may provide many valuable traits for wheat improvement (Cao *et al*. 2008, Du *et al*. 2014, Ma *et al*. 2016). There have been attempts to transfer alien Ns genome-encoded prolamins, in particular the HMM-prolamins of *Psathyrostachys*, into wheat by distant hybridization (Zhao *et al*. 2010). Currently, little is known about the HMM-prolamins that are encoded by the Ns genome of *Psathyrostachys*, although the results presented by Kong *et al*. (2014) suggest that the HMM-prolamins of *Ps. juncea* are very similar to wheat HMM-GSs. Four *Glu-Ns1* alleles that correspond to smaller subunits have been isolated, while the genes related to the larger subunits remain unknown. *Leymus* is a genus comprising ~30 species of polyploid perennial grasses that are closely related to the Ns genome of *Psathyrostachys*. Species within the *Leymus* genus share two basic Ns and Xm genomes (Zhang and Dvořák 1991). In an attempt to isolate HMM-prolamins encoded by the Ns genome from tetraploid *Leymus* species, three different HMM-GS genes were identified (Sun *et al*. 2014), and the small subunit open reading frames (ORFs) were cloned for further expression in bacteria. As the gene sequences of the larger subunits are not defined clearly, we sought to elucidate the gene sequences using PCR methods. We designed new oligo primers that allowed us to isolate and express new genes that encoded these proteins. This is the first report of newly identified D-hordein proteins from a species not belonging to *Hordeum*.

Materials and methods

Plants and DNA extraction: A total of 15 *Psathyrostachys juncea* (Fisch.) Nevski and *Ps. fragilis* (Boiss) Nevski accessions derived from nine different countries, which were kindly supplied by USDA-ARS, were used in this study (Table 1 Suppl.). Genomic DNA was extracted from 2-week-old seedlings using the cetyltrimethyl ammonium bromide (CTAB) extraction method described by Doyle and Doyle (1990).

SDS-PAGE analysis: HMM-prolamins from one (for all accessions) or six individual seeds (for *Ps. juncea* W6 17986, PI 434242, PI 531828, and PI 549118) were extracted from endosperms with an extraction buffer [0.0625 mM Tris-HCl, pH 6.8, 2 % (m/v) sodium dodecylsulphate (SDS), 5 % (v/v) β-mercaptoethanol, 10 % (v/v) glycerol, and 0.002 % (m/v) bromophenol blue] based on a ratio of 4 mg of sample to 0.1 cm^3 of extraction buffer. The mixtures were denatured in boiling water and centrifuged at 15 000 *g* for 5 min after gently shaking the samples at room temperature for 1 h. Next, 5 mm^3 of supernatant was loaded on a 10 % (m/v) vertical SDS-PAGE gels for separation of HMM-prolamins (Dai *et al*. 2013).

Isolation of HMM-prolamins encoding ORFs from *Ps. juncea*: Two forward (PF1 and PF2) and two reverse primers (PR1 and PR2) (Figs. 1*A* Suppl. and 2*A*,*B* Suppl., Table 2 Suppl.) were designed based on the conserved nucleotides corresponding to the ORF and its 5'- and 3'-UTR flanking sequences of the published HMM-GSs and D-hordeins. Three PCR primer combinations, PF1/PR1, PF2/PR1, and PF2/PR2 were used to amplify the full ORFs of the HMM-prolamins from *Ps. juncea* PI 315080 (Fig. 1*B* Suppl.). To confirm the natural in-frame stop codon in *Ns 1.3*, we re-amplified the ORFs of *Ns 1.3* using primers Ns1.3-F/Ns1.3-R (Table 2 Suppl.). High-fidelity *CloneAmp HiFi PCR Premix* (*Clontech*, Dalian, China) was used for all PCR reactions. The PCR amplification conditions were 95 ºC for 3 min, followed by 35 cycles at 98 ºC for 10 s, 67 ºC for 15 s, and 72 ºC for 4 min, with a final extension at 72 ºC for 5 min. The PCR products were separated on 1 % agarose

gels. The purified DNA fragments were cloned into the plasmid vector pJET1.2, and the resultant positive clones were selected after transforming into chemically-competent *Escherichia coli* JM109 cells. The positive plasmids containing inserts were purified with a *QIAprep® Spin Miniprep* kit (*Qiagen*, Mississauga, Canada), and used for sequencing. The full-length ORFs were acquired by primer walking. Five short sequences, each with ~800 bp in DNA length and 100 - 200 bp overlaps, were assembled into a full-length sequence by *DNAMAN 8.0*. Multiple sequence alignments of D-hordeins and HMM-GSs from various *Triticeae* species were completed by using the *Clustal W* program (Thompson *et al*. 1994).

Isolation of the N- and C-terminal sequences from two types of *Ps. juncea* D-hordeins: Four randomly selected accessions (W6 19786, PI 434242, PI 531828, and PI 549118) were used for cloning the N- and C-termini of the HMM-prolamins. Two different D-hordeins (small and large) were present in *Ps. juncea*. The differences between the two D-hordeins lie within the N-terminal (about 480 bp) and C-terminal (about 300 bp). PCR amplification of N- and C-terminal parts of the small hordein was performed using PF2/N-PR1 and C-PF1/C-PR1, and the large hordein using PF2/N-PR2 and C-PF2/C-PR2 (Fig. 2*C*,*D* Suppl.). The DNA fragments were cloned as mentioned above. At least six positive clones for each transformation event were sequenced in order to collect enough sequence data.

Bacterial expression of *Ns 2.6* and *Ns 2.9* ORFs: Two plasmid constructs, *pET-30a-Ns 2.6* and *pET-30a-Ns 2.9*, were used for bacterial expression of *Ps. juncea* Ns 2.6 and Ns 2.9 ORFs. Signal peptides were removed from the original ORF before insertion into the expression vector *pET-30a* (*Invitrogen*, Carlsbad, CA, USA). Two restriction enzyme sites, *Nde*I and *Eco*RI, were introduced into the recombinant plasmids by PCR mutagenesis. PCR primers used for bacterial plasmid construction were listed in Table 2 Suppl. Protein expression of *pET-30a-Ns 2.6* and *pET-30a-Ns 2.9* were conducted in *E. coli* strain BL21(DE3)plysS. Recombinant protein expression was induced by adding 1 mM isopropyl-β-D-1-galactoside (IPTG) once the cell concentration reached absorbance (A_{600}) 0.6 and then incubated at 37 ºC for 3 h with gently shaking. The recombinant proteins were analyzed by SDS-PAGE using protein extract from an uninduced bacterial culture as negative control.

Phylogenetic analysis: Phylogenetic trees were constructed using the N- and C-terminal sequences of D-hordeins from two *Hordeum* species (*H. vulgare* L. and *H. chilense* Roemmer et Schultes) in the GenBank database (https://www.ncbi.nlm.nih.gov/) and from *Ps. juncea* in this study (Tables 3 Suppl.). Multiple sequence alignments were conducted with *Clustal W* (Thompson *et al*. 1994). Two maximum likelihood trees of the conserved N- and C-terminal residues were established using the Jones-Taylor-Thornton model of *MEGA 5.0* following complete gap deletion (Tamura *et al*. 2011). The bootstrap values were set up to 1 000 replications.

The divergence times among D-hordeins from two *Hordeum* species and *Ps. Juncea* as well as the HMM-GSs from wheat and its wild relatives (Table 1), were estimated using the nucleotide sequences of the conserved N- plus C-terminal domains of the compared genes. In brief, the divergence times (t) among each pairs of aligned genes were calculated by linear regression of molecular distance (D) and one global rate of substitution (r) based on the molecular clock model using t = D/2r (see review by Rutschmann 2006). The mean distances among loci were computed by *MEGA 5.0* according to the conserved nucleotides of the N- plus C-terminal. The global rate of substitution was calculated using an average divergence times of 12.2 MYA between wheat and barley previously estimated by Huang *et al*. (2002) (11.4 MYA) and Gaut (2002) (13.0 MYA), and an average molecular distance of 0.192 between *Glu-1* of wheat and D-hordein loci of barley (Table 1), then got a global rate of substitution $r = 7.9 \times 10^{-9}$ per site per year.

Results

SDS-PAGE analysis revealed that individual seeds of each of the 13 *Ps. juncea* and 2 *Ps. fragilis* accessions expressed 1 - 2 HMM-prolamin subunits in the same molecular mass range as the HMM-GS of wheat reference cultivars Chinese Spring and Shinchunaga (Fig. 1*A*). Analysis of multiple seeds of the four *Ps. juncea* accessions revealed that they were heterozygous at the HMM-prolamins locus (Fig. 1*B*). The electrophoretic mobility of these subunits was slower than that of the HMM-GS Dx2, but similar to that of Dx2.2, as observed in the two wheat controls (Fig. 1*A*). Unexpectedly, the smaller subunits that were closer to the wheat Dy12 described by Kong *et al*. (2014) were not observed in this study.

Newly designed primers facilitated amplification of DNA fragments related to the larger subunits (Figs. 1*A* Suppl. and 2*A*,*B* Suppl.). A total of seven fragments were amplified using the three primer combinations, and three of the amplicons were verified to contain HMM-prolamin ORFs (Fig. 1*B* Suppl.). In detail, two fragments, 1.3 and 2.0 kb, were amplified using the P1 primer combination (PF1+PR1), and the 1.3 kb band was confirmed to contain HMM-prolamin ORFs, while the 2.0 kb fragment corresponded to a nonspecific amplification. One ~2.6 kb fragment was amplified using the P2 primer combination (PF2+PR2), and confirmed to be an HMM-prolamin.

Four different amplicons were obtained using the P3 primer combination (PF2+PR1), and only the largest 2.9 kb fragment was cloned and verified; the remaining three bands were not cloned for further study (Fig. 1*B*

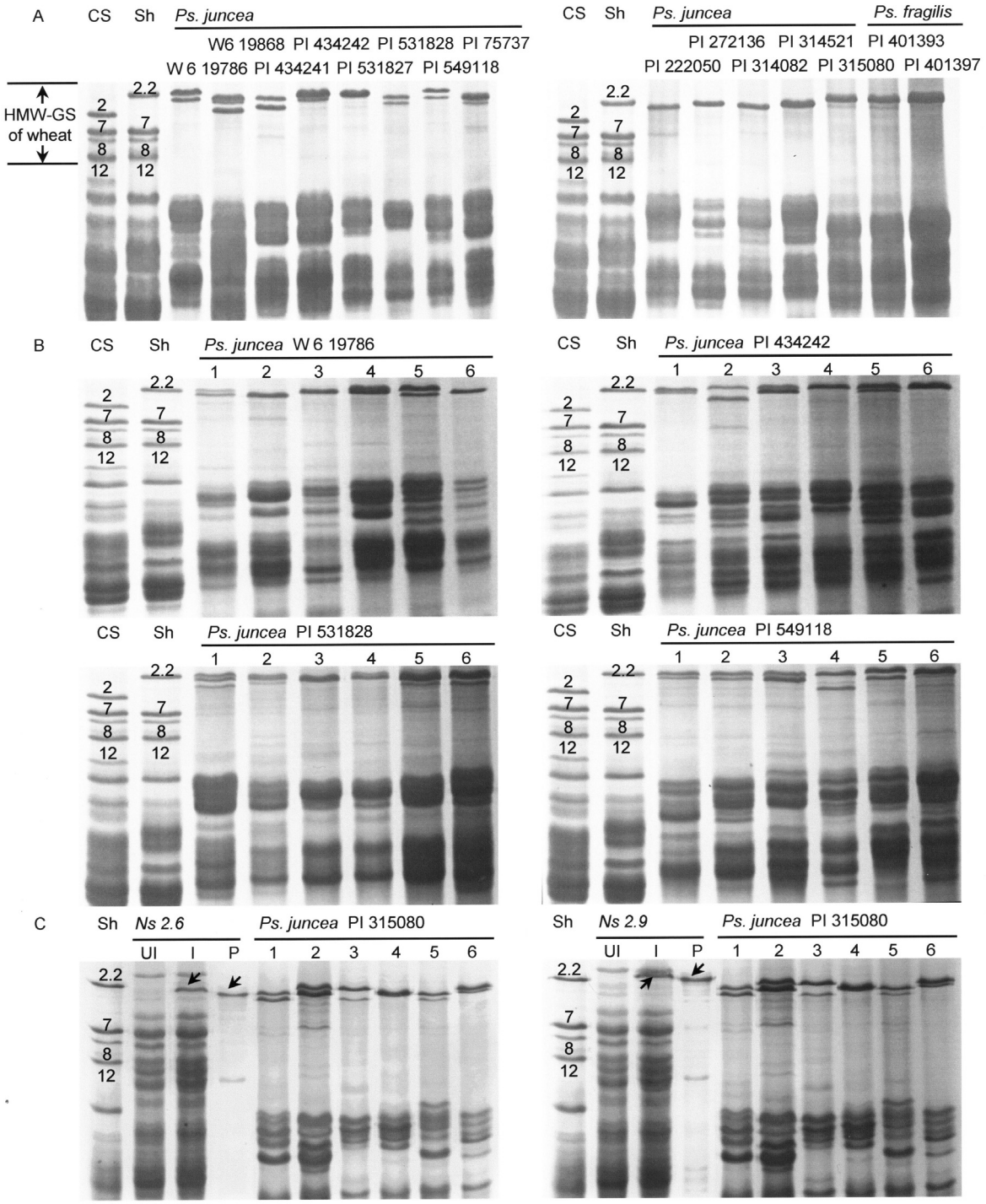

Fig. 1. SDS-PAGE analysis of HMM-prolamins from *Psathyrostachys juncea* and *Ps. fragilis*. *A* - Rough analysis of HMM-prolamins from single seed of the 13 *Ps. juncea* and 2 *Ps. fragilis* accessions. *B* - Detailed analysis of HMM-prolamins for multiple seeds of four *Ps. juncea* accessions. *C* - Bacterial expression of *Ps. juncea* HMM-prolamins Ns 2.6 (*left*) and Ns 2.9 (*right*) in *Escherichia coli* BL21(DE3)plysS; the expressed proteins are indicated by *arrowheads* in lanes I and P. Bacterial cultures are induced by 1 mM IPTG (I) and purified by selective precipitation with 50 % (v/v) propanol plus 2 % (m/v) DTT (P). Lane UI is from bacterial culture without being induced by IPTG. Wheat cultivars CS (Chinese Spring) and Sh (Shinchunaga) were used as wheat seed extract controls. Numbers in the CS and Sh lanes indicate the position of known HMM-GS for two wheat controls.

Suppl.). The genes corresponding to the 1.3, 2.6, and 2.9 kb bands were named *Ns 1.3*, *Ns 2.6* and *Ns 2.9*, respectively. A primer walking method was used to obtain the full nucleotide sequences of *Ns 2.6* and *Ns 2.9*. The total lengths of *Ns 1.3*, *Ns 2.6*, and *Ns 2.9* were 1 359, 2 788, and 3 004 bp, respectively, and their respective ORFs were 1 320, 2 628, and 2 901 bp after excluding the 5'-upstream and 3'-downstream partial sequences. The GenBank accession numbers for them are KY587711 - KY587713.

Three ORFs, *Ns 1.3*, *Ns 2.6* and *Ns 2.9*, were translated into amino acid sequences using the universal triplet genetic code, producing two mature proteins (without signal peptides) for *Ns 2.6* and *Ns 2.9* with 853 and 944 amino acid (AA) residues, respectively. *Ps. juncea* Ns 1.3, Ns 2.6, and Ns 2.9 shared similar structural characteristics with D-hordeins of *H. chilense* and *H. vulgare*. For example, they all contained a large central repetitive domain including three sub-domains: the N-terminal repeats (R1), the central degenerate repeats (R2), and the C-terminal repeats (R3) (Fig. 2, Fig. 3 Suppl., Table 3 Suppl.). It strongly demonstrates that the HMM-prolamins of *Ps. juncea* belong to the D-hordein family. Meanwhile, *Ns 2.9* encodes a unique mature protein with 944 AA residues, which has been identified as the largest D-hordein currently reported. The D-hordeins of *Ps. juncea* were very similar to those from *H. chilense*, but with minor modifications to *H. vulgare*. Both the D-hordeins from *Ps. juncea* and *H. chilense* had a total of 9 cysteine residues, including one each in the R2 and R3 sub-domains of their repetitive domains, and a shorter R3 (~70 AA), while in D-hordeins of *H. vulgare*, there were a total of 10 cysteine residues, including 2 and 2 cysteine residues in the R2 and R3 repetitive domains, respectively, and a longer R3 (~180 AA).

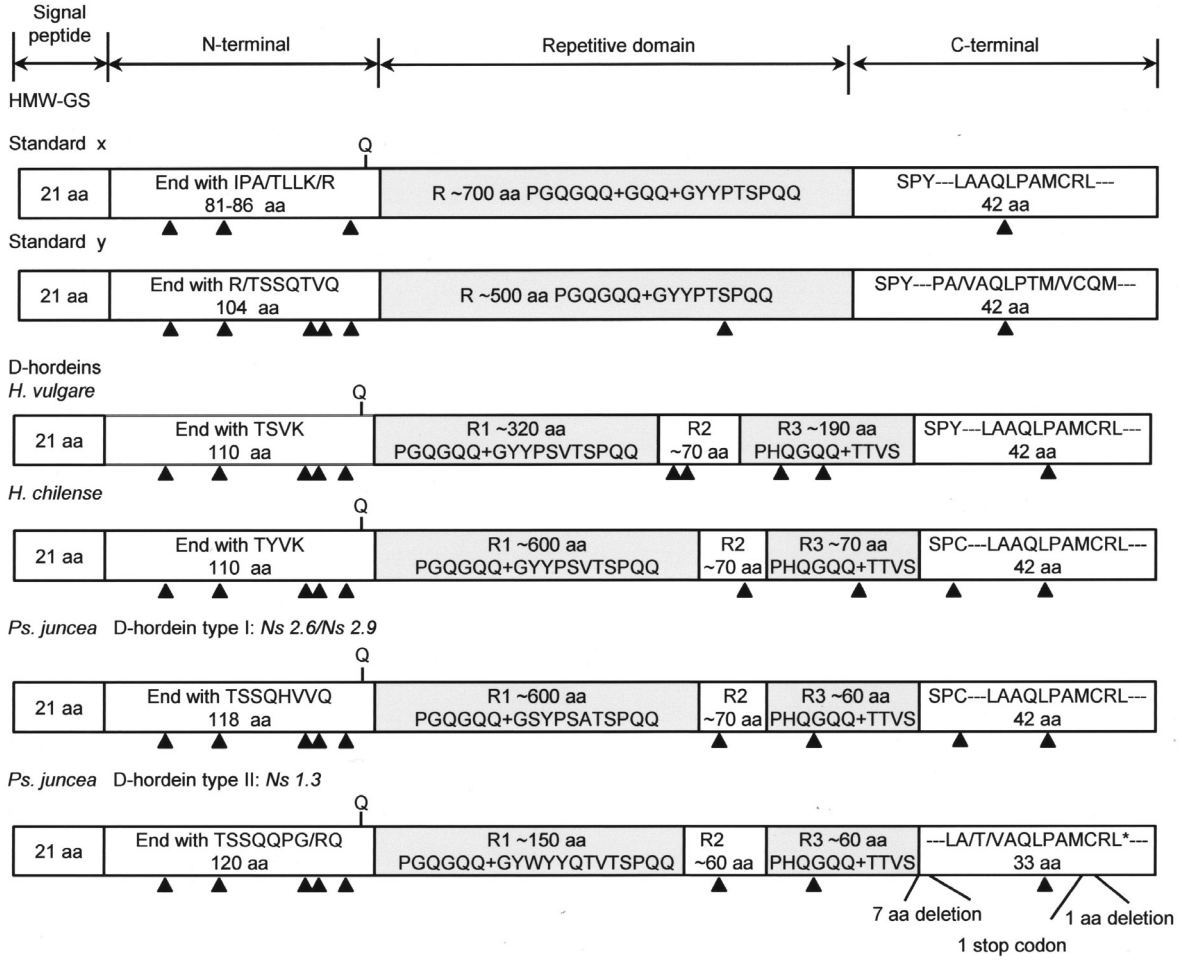

Fig. 2. Schematic representation of D-hordeins from *Hordeum vulgare*, *H. chilense*, and *Psathyrostachys juncea*, as well as HMM-GSs from wheat.

Multiple sequence alignments of the D-hordeins from *Ps. juncea*, *H. chilense*, and *H. vulgare* revealed that the *Ps. juncea* D-hordeins contained a novel feature. The *Ps. juncea* D-hordeins had longer N-terminal domains (118 - 120 AA residues) than those from *Hordeum* species (*H. chilense* and *H. vulgare*, both with 110 AA residues) (Figs. 2 and 3). The three D-hordeins from *Ps. juncea* could be further divided into two types, Ns 1.3

and Ns 2.6/Ns 2.9, and significant differences could be observed between the two D-hordein types (Fig. 2). Firstly, the two types had different ORF lengths: Ns 1.3 was < 1.5 kb and Ns 2.6/Ns 2.9 were > 2.5 kb. Secondly, the motif units in the repetitive domain R1 were different. The Ns 1.3 motif units were tridecapeptide (consensus GYWYYQTVTSPQQ), whereas they were undeca-peptide (consensus GSYPSATSPQQ) for Ns 2.6/Ns 2.9 (Fig. 3 Suppl.). Third, Ns 1.3 had a shorter C-terminal length than Ns 2.6 and Ns 2.9, and contained a premature stop codon and a unique C-terminus (with a modified short MLANE peptide instead of the normal LSASQ C-terminal for D-hordeins and HMM-GSs) (Fig. 3, Fig. 4*A*,*B* Suppl.). Together, this suggested that Ns 1.3 was likely not expressed as a functional protein, while Ns 2.6 and Ns 2.9 were expected to be functionally expressed proteins.

The ORFs corresponding to the *Ps. juncea* D-hordeins of *Ns 2.6* and *Ns 2.9* were expressed in bacteria without the signal peptide (Fig. 1*C*). SDS-PAGE analysis revealed that the ORFs of *Ns 2.6* and *Ns 2.9* were expressed in *E. coli* BL21(DE3)plysS following induction by IPTG (Fig. 1*C*) in comparison to bacterial culture extracts without being induced by IPTG (Fig. 1*C*). Target proteins were also detected in the bacterial products selectively extracted with 50 % (v/v) propanol including 2 % (m/v) dithiotreitol (DTT) to exclude *E. coli* proteins (Fig. 1*C*). The electrophoretic mobility of recombinant proteins was very close to native proteins from seed extracts and wheat reference. However, Ns 2.6 and Ns 2.9 proteins expressed in bacteria did not clearly correspond to the HMM-prolamins of *Ps. juncea* PI 315080, as the D-hordein locus was heterozygous (Fig. 1*C*).

All of the sequences determined in this project were used for phylogenetic analyses and were stored in the GenBank database using accession numbers KY587660 to KY587713. Two maximum likelihood trees were constructed using the AA residues of the N- and C-termini, respectively. In the topology tree based on the N-terminal AA residues, all of the genes were divided into two distinct clades based on a very high bootstrap value (Fig. 4*A*). *Ns 1.3* and its allelic variants formed a separate clade (*Ps. juncea* clade II) that was in parallel to the clade (clade I) formed by *Ns 2.6/Ns 2.9*, their allelic variants (*Ps. juncea* clade I), and D-hordeins from *H. chilense* and *H. vulgare.* Within clade I, the D-hordeins from *H. vulgare* were separated from *H. chilense* and *Ps. juncea* clade I (*Ns 2.6/Ns 2.9* and their allelic variants) by strong bootstrap support.

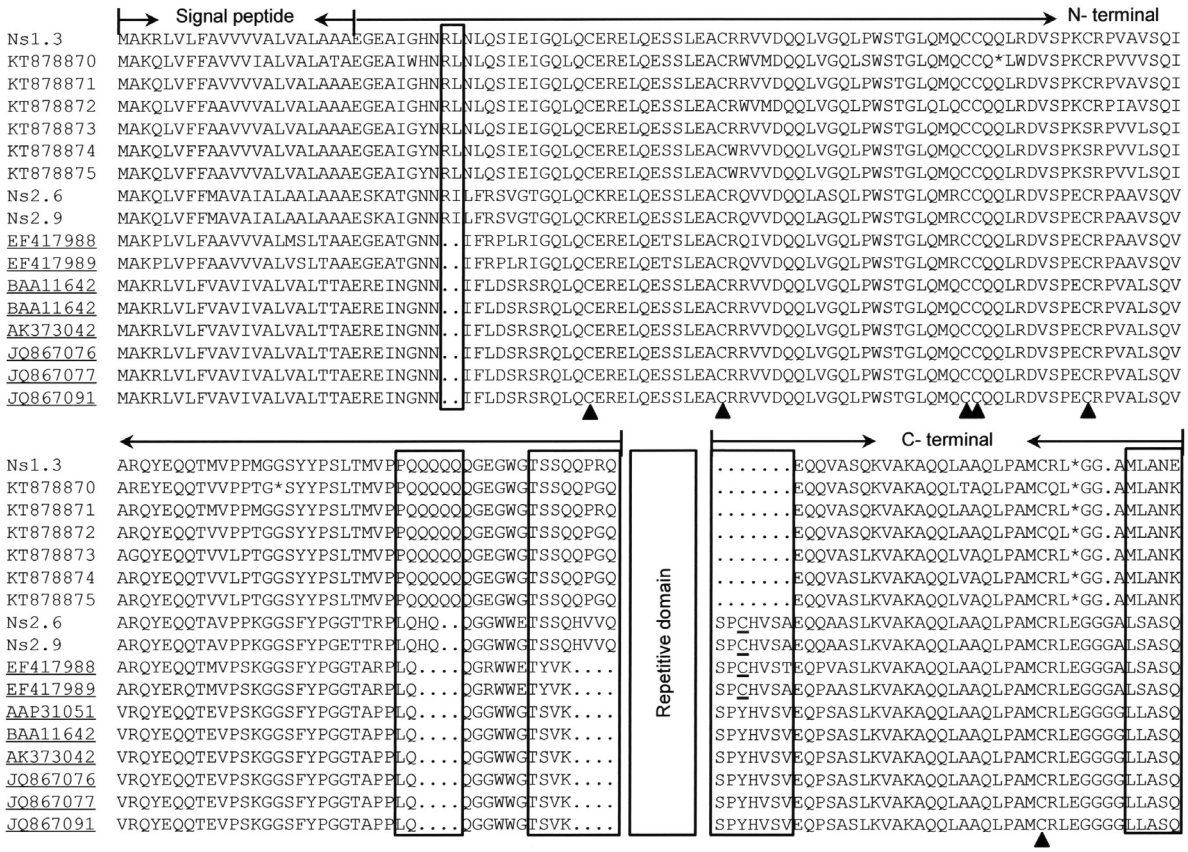

Fig. 3. Multiple sequence alignments of D-hordeins from *Hordeum vulgare*, *H. chilense*, and *Psathyrostachys juncea*. The sequences from *Hordeum* species (*H. vulgare* and *H. chilense*) are *underlined*, and unpublished data (KT 878870 - KT 878875) are given a KT prefix. The main differences among D-hordeins are *boxed*. The conserved cysteines are marked by *arrowheads* and the substituted cysteines are *underlined. Asterisks* indicate the in-frame stop codons.

Although D-hordeins from *H. chilense* and *Ps. juncea* clade I were very similar, they were still clearly separated (Fig. 4*A*). The tree constructed by the C-terminal AA residues resulted in a tree similar to the N-terminal analysis (Fig. 4*B*). As expected, the D-hordeins from *Ps. juncea* in both trees were divided into two different branches: *Ps. juncea* clade I (*Ns 2.6/Ns 2.9* and their allelic variants) and *Ps. juncea* clade II (*Ns 1.3* and its allelic variants), with *Ps. juncea* clade II distantly related to the D-hordeins from *H. chilense* and *H. vulgare*, and *Ps. juncea* clade I closely related to D-hordeins of *H. chilense* (Fig. 4*B*).

The divergence times among the D-hordeins and HMM-GSs loci of *Triticeae* species were calculated (Tables 1 and 3 Suppl.). In order to avoid overestimation of the divergence times, the two types of D-hordeins from *Ps. juncea* were calculated separately with the orthologous D-hordein loci of *Hordeum* species and the HMM-GS from wheat and related wild species, as the phylogenetic analysis supported the suppositions that they belonged to two independent clades. The divergence times among the D-hordeins from *H. vulgare*, *H. chilense*, and *Ns 2.6/Ns 2.9* of *Ps. juncea* with orthologous *Glu-1* from wheat and other relative species, ranging from 6.6 to 13.6 million years ago (MYA), were shorter than those obtained for *Ns 1.3* of *Ps. juncea* (varying from 14.6 to 20.8 MYA) with each orthologous locus (Table 1). Combining the *Ns 1.3* and *Ns 2.6/Ns 2.9* as one loci, the divergence times for *Ns 1.3* with orthologous loci of wheat and other relative species were underestimated by 1.5 - 2.0 MYA, with an average of about 1.6 MYA, but *Ns 2.6/Ns 2.9* were overestimated by 5.1 - 6.2 MYA, with a mean of about 5.8 MYA (Table 1). *Ns 1.3* diverged earlier (>14 MYA) from *Triticeae* than the other three D-hordeins (*H. vulgare*, *H. chilense*, *Ns 2.6/Ns 2.9*). However, *Ns 2.6/Ns 2.9* diverged more recently from the D-hordeins of *H. chilense* and *H. vulgare*, at 5.4 and 7.8 MYAs, respectively. Further, the divergence times between the D-hordeins of *H. chilense* and *Ns 2.6/Ns 2.9* of *Ps. juncea* (5.4 MYA) were shorter than the separation between the D-hordeins from *H. vulgare* and *H. chilense* (7.3 MYA).

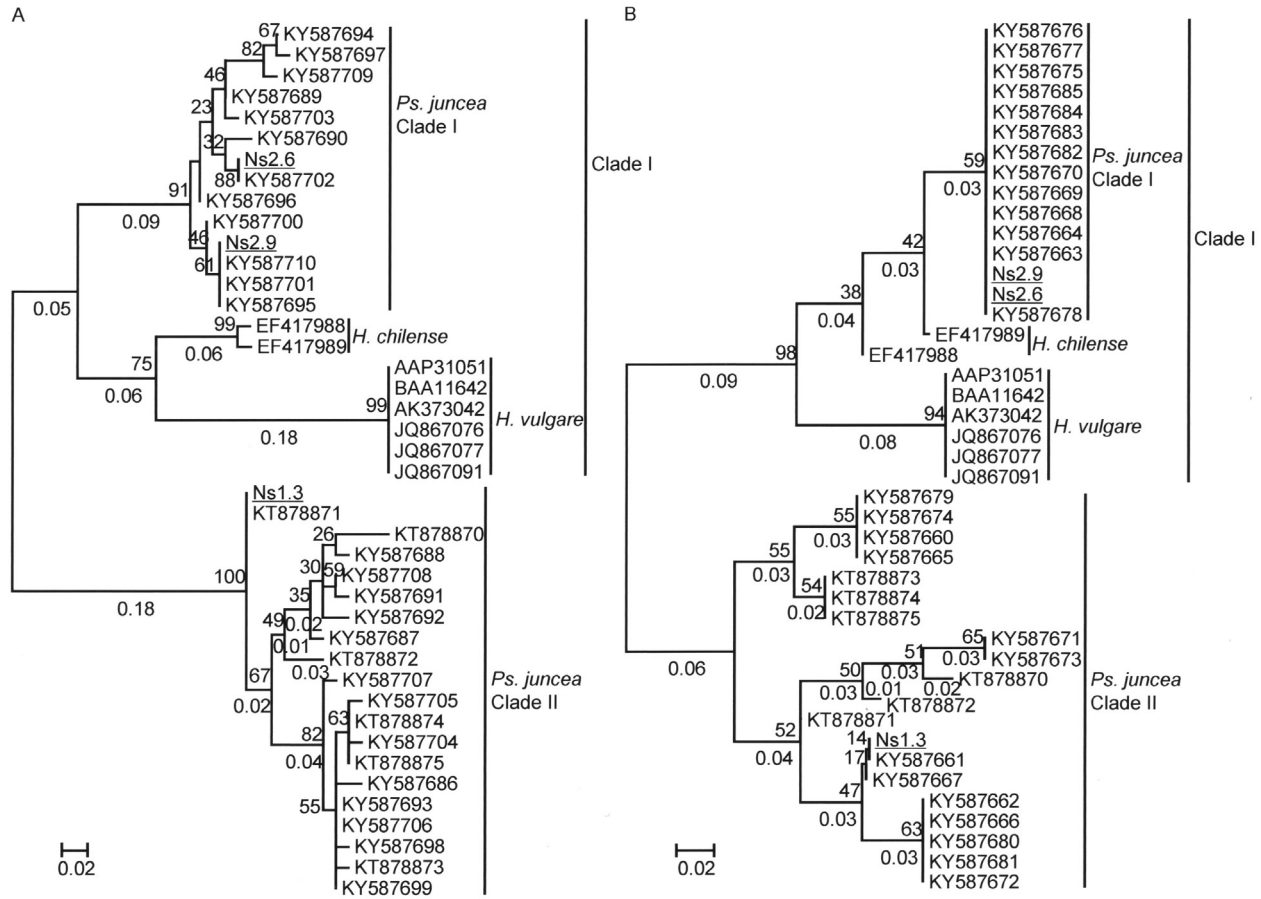

Fig. 4. Phylogenetic analysis of D-hordeins from *Psathyrostachys juncea*, *Hordeum chilense*, and *H. vulgare* based on the N- (*A*) and C-terminal (*B*) amino acid residues, respectively. Bootstrap values are calculated using 1 000 replications and are shown above the lines. Branch lengths are shown under the lines. The three D-hordeins from *Ps. juncea* PI 315080 are *underlined*.

Table 1. Divergence times among D-hordeins from two *Hordeum* species and *Ps. juncea* and HMM-GS orthologous loci from wheat and its relatives.

Loci	*H. vulgare*, Hv *D-hordein* distance	MYA	*H. chinense*, Hc *D-hordein* distance	MYA	*Ps. juncea Ns2.6/Ns2.9* type distance	MYA	*Ps. juncea Ns1.3* type distance	MYA	*Ps. juncea Ns1.3/Ns2.6/Ns2.9* distance	MYA
Ns1.3/Ns2.6/Ns2.9	0.210	13.3	0.206	13.0	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Ns1.3 type	0.235	14.9	0.240	15.2	0.222	14.1	n.a.	n.a.	n.a.	n.a.
Ns2.6/Ns2.9 type	0.123	7.8	0.086	5.4	n.a.	n.a.	0.222	14.1	n.a.	n.a.
D-hordein-Hc	0.115	7.3	n.a.	n.a.	0.086	5.4	0.240	15.2	0.206	13.0
D-hordein-Hv	n.a.	n.a.	0.115	7.3	0.123	7.8	0.235	14.9	0.210	13.3
Glu-St1	0.116	7.3	0.106	6.7	0.104	6.6	0.230	14.6	0.202	12.8
Glu-F1	0.118	7.5	0.120	7.6	0.113	7.2	0.240	15.2	0.212	13.4
Glu-O1	0.126	8.0	0.111	7.0	0.105	6.6	0.250	15.8	0.218	13.8
Glu-Q1	0.130	8.2	0.120	7.6	0.123	7.8	0.252	15.9	0.224	14.2
Glu-Ns1	0.134	8.5	0.135	8.5	0.130	8.2	0.244	15.4	0.219	13.9
Glu-Xe1	0.146	9.2	0.142	9.0	0.144	9.1	0.249	15.8	0.226	14.3
Glu-W1	0.159	10.1	0.148	9.4	0.144	9.1	0.262	16.6	0.236	14.9
Glu-K1	0.172	10.9	0.174	11.0	0.175	11.1	0.285	18.0	0.260	16.5
Glu-E1	0.175	11.1	0.184	11.6	0.178	11.3	0.286	18.1	0.262	16.6
Glu-U1	0.182	11.5	0.164	10.4	0.179	11.3	0.293	18.5	0.268	17.0
Glu-1	0.192	12.2	0.192	12.2	0.190	12.0	0.303	19.2	0.278	17.6
Glu-C1	0.194	12.3	0.172	10.9	0.182	11.5	0.292	18.5	0.268	17.0
Glu-Ta1	0.197	12.5	0.182	11.5	0.191	12.1	0.296	18.7	0.272	17.2
Glu-R1	0.215	13.6	0.211	13.4	0.212	13.4	0.328	20.8	0.302	19.1

Discussion

Previous studies have attempted to discover novel HMM-prolamin genes encoded by the Ns genomes of *Psathyrostachys* and *Leymus* species (Kong *et al*. 2014, Sun *et al*. 2014). SDS-PAGE analysis of the HMM-prolamins from the species of both genera identified a larger type of subunit that exhibited slower electrophoretic mobility than HMM-GS Ax1 and Ax2*, as well as a smaller type that was faster than wheat HMM-GS Dy12 (Kong *et al*. 2014, Sun *et al*. 2014). The genes encoding the smaller subunits from species of both genera were isolated and expressed in bacteria, but the genes related to the larger subunits were still not further characterized. In this study, three HMM-prolamins (Ns 1.3, Ns 2.6, and Ns 2.9) were isolated from *Ps. juncea* PI 315080, their amino acid sequences confirmed that they were closely related to the larger subunit, and bacterial expression showed that the ORFs for *Ns 2.6* and *Ns 2.9* could be normally expressed. The three *Ps. juncea* HMM-prolamins shared four similar structural domains with D-hordeins from two *Hordeum* species, especially a unique larger central repetitive domain that consisted of three sub-domains including the N-terminal repeats (R1), the central degenerate repeats (R2), and the C-terminal repeats (R3), unlike the solely repetitive domain (R) of wheat HMM-GS (Gu *et al*. 2003, Shewry *et al*. 2003, Pistón *et al*. 2007). The tetrapeptide motif unit TTVS, which is specified by the C-terminal repeats (R3) of barley D-hordeins and has no counterpart in HMM-GS, were also present in Ns 1.3, Ns 2.6, and Ns 2.9 (Pistón *et al*. 2007, Fig. 3 Suppl.). Although the N-terminal lengths of the three *Ps. juncea* HMM-prolamins (118 - 120 residues) were longer than those from barley (110 residues) and the wheat HMM-GS (81 - 86 and 104 residues for x- and y-types, respectively), the mature proteins more closely resembled the D-hordeins of *Hordeum* species in both the sequences and compositions of the amino acid residues, as well as the occurrence and distributions of cysteine residues (Gu *et al*. 2003, Shewry *et al*. 2003, Pistón *et al*. 2007, Figs. 2, 3 and 3 Suppl.). Therefore, the three HMM-prolamins identified from *Ps. juncea* were actually D-hordeins. Previously, D-hordeins have been identified from *Hordeum* species, as well as the derivate cereal *Tritordeum* (Alvarez *et al*. 1999). Three D-hordeins identified from *Ps. juncea* suggest that *Psathyrostachys* species may be another source of D-hordeins.

The D-hordeins of the *Ps. juncea* Ns 2.6/Ns 2.9 type were very similar to those in *H. chilense*, both of which were larger than the D-hordeins of *H. vulgare* and the HMM-GS of wheat due to the occurrence of longer repetitive domains. The D-hordeins from *Ps. juncea*, *H. vulgare*, and *H. chilense* differed from wheat HMM-GS by having more cysteine residues distributed in a different manner. The repetitive domain length, as well as the numbers and distributions of cysteine residues, are two structural features of wheat HMM-GS that highly contribute to dough strength (Tamás *et al*. 2002). Subunits with longer and highly regular repetitive domains must become more elastic due to the existence of more interchain hydrogen bonds between the subunits

(Gilbert *et al*. 2000) or form more β-turns (Tatham *et al*. 1985) within repetitive domains, and could form more stable interactions during the mixing process. The above theories were verified *via* experiments using spectroscopic and biomechanical treatment of gluten fractions and recombinant proteins (Feeney *et al*. 2003, Wellner *et al*. 2005, 2006). Interchain disulphide bonds between cysteine residues are also considered to be major factors affecting the stability of glutenin polymers (Shewry *et al*. 2003). Usually, the x-type HMM-GSs have a total of four cysteine residues, while the y-type have seven cysteine residues present in mature proteins, with respectively three and five located in the N-terminus, and both with one at the C-terminus. In contrast, the D-hordeins from *H. chilense* had nine cysteine residues, with five, two, and two respectively distributed in the N- and C-terminus, and repetitive domains, while those from *H. vulgare* have ten cysteine residues, with five, one, and four localized to above corresponding domains (Gu *et al*. 2003, Pistón *et al*. 2007, Table 3 Suppl.).

Wheat lines harboring HMM-GS pairs 5+10 show higher ratios of large glutenin polymers than their sister lines carrying allelic pairs of 2+12, as explained by the extra cysteine residue located at the beginning of the repetitive domain in Dx5 (Popineau *et al*. 1994). The D-hordeins of *Ps. juncea* and *H. chilense* differed from wheat HMM-GS by carrying more cysteine residues, and were distinguished from those of *H. vulgare* by odd numbers of cysteine residues instead of even numbers, which would be favorable in forming longer and strengthened gluten polymers, as the odd numbers of cysteine residues tend to form more inter-molecular disulphide bonds between subunits, whereas even numbers of cysteine residues form more intra-molecular disulphide bonds within subunits. Importantly, longer repetitive domain would benefit from forming extended gluten polymers in comparison to proteins with shorter repetitive domains. The D-hordeins of *H. chilense* were reported to play an important role in regulating the bread-making and malting quality in the artificially developed new cereal, *Tritordeum* (Alvarez *et al*. 1992, Martín *et al*. 1999), and the contribution ratios of D-hordein variations from *H. chilense* parent to the dough strength were verified after excluding the effects contributed by the HMM-GSs of the durum wheat parent (Alvarez *et al*. 1999). The D-hordeins from *H. chilense* and *Ps. juncea* showed higher similarities in the repetitive domain organization and in the numbers and distributions of cysteine residues when compared to the D-hordeins from *H. vulgare* and the HMM-GSs from wheat, possibly suggesting that they may have the ability to form more stable interactions between subunits than those formed by D-hordeins of *H. vulgare* and the HMM-GSs of wheat. Novel D-hordeins identified from *Ps. juncea* would probably change the end-use quality of wheat flours if transferred into a wheat background. It also should be noted that some studies have tried introducing the Ns genome-encoded HMM-prolamins into wheat through distant hybridization method (Zhao *et al*. 2010).

Three novel D-hordeins, Ns 1.3, Ns 2.6 and Ns 2.9, were characterized from *Ps. juncea*. Up to now, D-hordeins have been identified only in the species of the genus *Hordeum*. The present results extended the distribution of D-hordeins to genus *Psathyrostachys*, and expanded the source of D-hordeins beyond *Hordeum*. The *Ps. juncea* D-hordeins were divided into two types, larger type I (Ns 2.6/Ns 2.9 and their allelic variants) and smaller type II (Ns 1.3 and its allelic variants), based on the lengths of ORFs for the genes, the motif units in repetitive domains, and phylogenetic analysis. The *Ps. juncea* D-hordeins provide new information for understanding of the genes and evolution of HMM-prolamins in *Triticeae* species, and also identify a novel type of D-hordeins that have a potential utilization in altering the biochemical function of wheat flour.

References

Alvarez, J.B., Ballesteros, J., Sillero, J.A., Martín, L.M.: *Tritordeum*: a new crop of potential importance in the food industry. - Hereditas **116**: 193-197, 1992.

Alvarez, J.B., Campos, L.A.C., Martín, A., Sillero, J.A., Martín, L.M.: Genetic analysis of prolamins synthesised by the H^{ch} genome and their effects on gluten strength in hexaploid *Tritordeum*. - Euphytica **107**: 177-184, 1999.

Atanassov, P., Borries, C., Zaharieva, M., Monneveux, P.: Hordein polymorphism and variation of agromorphological traits in a collection of naked barley. - Genet. Resour. Crop Evol. **48**: 353-360, 2001.

Baden, C., Bothmer, R., Flink, J., Jacobsen, N.: Intergeneric hybridization between *Psathyrostachys* and *Hordeum*. - Nord. J. Bot. **9**: 333-342, 1989.

Baden, C.: A taxonomic revision of *Psathyrostachys* (*Poaceae*). - Nord. J. Bot. **11**: 3-26, 1991.

Blake, T.K., Ullrich, S.E., Nilan, R.A.: Mapping of the *Hor-3* locus encoding D hordein in barley. - Theor. appl. Genet. **63**: 367-371, 1982.

Cao, Z., Deng, Z., Wang, M., Wang, X., Jing, J., Zhang, X., Shang, H., Li, Z.: Inheritance and molecular mapping of an alien stripe-rust resistance gene from a wheat-*Psathyrostachys huashanica* translocation line. - Plant Sci. **174**: 544-549, 2008.

Dai, S.F., Pu, Z.J., Liu, D.C., Wei, Y.M., Zheng, Y.L., Hu, X.K., Yan, Z.H.: Characterization of novel HMW-GS in two diploid species of *Eremopyrum*. - Gene **519**: 55-59, 2013.

Doyle, J.J., Doyle, J.L.: Isolation of plant DNA from fresh tissue. - Focus **12**: 13-15, 1990.

Du, W.L., Wang, J., Lu, M., Sun, S.G., Chen, X.H., Zhao, J.X., Yang, Q.H., Wu, J.: Characterization of a wheat-*Psathyrostachys huashanica* Keng 4Ns disomic addition line for enhanced tiller numbers and stripe rust resistance. - Planta **239**: 97-105, 2014.

Feeney, K.A., Wellner, N., Gilbert, S.M., Halford, N.G., Tatham, A.S., Shewry, P.R., Belton, P.S.: Molecular structures and interactions of repetitive peptides based on wheat glutenin subunits depend on chain length. -

Biopolymers **72**: 123-131, 2003.
Gaut, B.S.: Evolutionary dynamics of grass genomes. - New Phytol. **154**: 15-28, 2002.
Gilbert, S.M., Wellner, N., Belton, P.S., Greenfield, J.A., Siligardic, G., Shewry, P.R., Tatham, A.S.: Expression and characterization of a highly repetitive peptide derived from a wheat seed storage protein. - BBA **1479**: 135-146, 2000.
Gu, Y.Q., Anderson, O.D., Londeore, C.F., Kong, X., Chibbar, R.N., Lazo, G.R.: Structural organization of the barley D-hordein locus in comparison with its orthologous regions of wheat genomes. - Genome **46**: 1084-1097, 2003.
Halford, N.G., Tatham, A.S., Sui, E., Daroda, L., Dreyer, T., Shewry, P.R.: Identification of a novel beta-turn-rich repeat motif in the D-hordeins of barley. - Biochim. biophys. Acta **1122**: 118-122, 1992.
Hsiao, C., Wang, R.R.C., Dewey, D.R.: Karyotype analysis and genome relationships of 22 diploid species in the tribe *Triticeae*. - Can. J. Genet. Cytol. **28**: 109-120, 1986.
Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R., Gornicki, P.: Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/ Aegilops* complex and the evolutionary history of polyploid wheat. - Proc. nat. Acad. Sci. USA **99**: 8133-8138, 2002.
Kong, L.N., Liang, Y., Qin, L.M., Sun, L., Xia, G.M., Liu, S.W.: Characterization of high molecular weight glutenin subunit genes from the Ns genome of *Psathyrostachys juncea*. - Dev. Genes Evol. **224**: 189-196, 2014.
Liu, S.W., Zhao, F., Gao, X., Chen, F.G., Xia, G.M.: A novel high molecular weight glutenin subunit from *Australopyrum retrofractum*. - Amino Acids **39**: 385-392, 2010.
Ma, D.F., Fang, Z.W., Yin, J.L., Chao, K.X., Jing, J.X., Li, Q., Wang, B.T.: Molecular mapping of stripe rust resistance gene *YrHu* derived from *Psathyrostachys huashanica*. - Mol. Breed. **36**: 64, 2016.
Martín, A., Alvarez, J.B., Martín, L.M., Barro, F., Ballesteros, J.: The development of *Tritordeum*: a novel cereal for food processing. - J. Cereal Sci. **30**: 85-95, 1999.
Payne, P.I.: Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. - Plant Biol. **38**: 141-153, 1987.
Pistón, F., Shewry, P.R., Barro, F.: D hordeins of *Hordeum chilense*: a novel source of variation for improvement of wheat. - Theor. appl. Genet. **115**: 77-86, 2007.
Popineau, Y., Cornec, M., Lefebvre, J., Marchylo, B.: Influence of high Mr glutenin subunits on glutenin polymers and rheological properties of gluten and gluten subfractions of near-isogenic lines of wheat Sicco. - J. Cereal Sci. **19**: 231-241, 1994.
Rasheed, A., Xia, X.C., Yan, Y.M., Appels, R., Mahmood, T., He, Z.H.: Wheat seed storage proteins: Advances in molecular genetics, diversity and breeding applications. - J. Cereal Sci. **60**: 11-24, 2014.
Rutschmann, F.: Molecular dating of phylogenetic trees: a brief review of current methods that estimate divergence times. - Divers. Distribut. **12**: 35-48, 2006.
Shewry, P.R., Halford, N.G.: Cereal seed storage proteins: structures, properties and role in grain utilization. - J. exp. Bot. **53**: 947-958, 2002.
Shewry, P.R., Halford, N.G., Tatham, A.S., Popineau, Y., Lafiandra, D., Belton, P.S.: The high molecular weight subunits of wheat glutenin and their role in determining wheat processing properties. - Adv. Food Nutr. Res. **45**: 219-302, 2003.
Shewry, P.R., Tatham, A.S.: The prolamin storage proteins of cereal seeds: structure and evolution. - Biochem. J. **267**: 1-12, 1990.
Sun, Y.Q., Pu, Z.J., Dai, S.F., Pu, X.X., Liu, D.C., Wu, B.H., Lan, X.J., Wei, Y.M., Zheng, Y.L., Yan, Z.H.: Characterization of y-type high-molecular-weight glutenins in tetraploid species of *Leymus*. - Dev. Genes Evol. **224**: 57-64, 2014.
Tamás, L., Gras, P.W., Solomon, R.G., Morell, M.K., Appels, R., Békés, F.: Chain extension and termination as a function of cysteine content and the length of the central repetitive domain in storage proteins. - J. Cereal Sci. **36**: 313-325, 2002.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S.: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. - Mol. Biol. Evol. **28**: 2731-2739, 2011.
Tatham, A.S., Miflin, B.J., Shewry, P.R.: The beta-turn conformation in wheat gluten proteins: relationship to gluten elasticity. - Cereal Chem. **62**: 405-412, 1985.
Thompson, J.D., Higgins, D.G., Gibson, T.J.: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. - Nucl. Acids Res. **22**: 4673-4680, 1994.
Wan, Y.F., Wang, D.W., Shewry, P.R., Halford, N.G.: Isolation and characterization of five novel high molecular weight subunit of glutenin genes from *Triticum timopheevii* and *Aegilops cylindrica*. - Theor. appl. Genet. **104**: 828-839, 2002.
Wellner, N., Marsh, J., Savag, A., Halford, N.G., Shewry, P.R., Mills, E.N.C., Belton, P.S.: Comparison of repetitive sequences derived from high molecular weight subunits of wheat glutenin, an elastomeric plant protein. - Biomacromolecules **7**: 1096-1103, 2006.
Wellner, N., Mills, E.N.C., Brownsey, G., Wilson, R.H., Brown, N., Freeman, J., Halford, N.G., Shewry, P.R., Belton, P.S.: Changes in protein secondary structure during gluten deformation studied by dynamic fourier transform infrared spectroscopy. - Biomacromolecules **6**: 255-261, 2005.
Yin, Y.Q., Ma, D.Q., Ding, Y.: Analysis of genetic diversity of hordein in wild close relatives of barley from Tibet. - Theor. appl. Genet. **107**: 837-842, 2003.
Zhang, H.B., Dvořák, J.: The genome origin of tetraploid species of *Leymus* (Poaceae: Triticeae) inferred from variation in repeated nucleotides sequences. - Amer. J. Bot. **78**: 871-884, 1991.
Zhang, L., Li, Z., Fan, R., Wei, B., Zhang, X.: Structural characterization and evolutionary relationship of high-molecular-weight glutenin subunit genes in *Roegnerianakaii* and *Roegneria alashanica*. - Int. J. mol. Sci. **17**: 1115, 2016.
Zhao, J.X., Ji, W.Q., Wu, J., Chen, X.H., Cheng, X.N., Wang, J.W., Pang, Y.H., Liu, S.H., Yang, Q.H.: Development and identification of a wheat-*Psathyrostachys huashanica* addition line carrying HMW-GS, LMW-GS and gliadin genes. - Genet. Resour. Crop Evol. **57**: 387-394, 2010.