RESEARCH NOTE

Identification of 40k γ-secalin genes

PENG-FEI QI[1]*, QING CHEN[1], THÉRÈSE OUELLET[2], CHENG-XING LE[1], JIA AI[1], YU-MING WEI[1], JI-RUI WANG[1] and YOU-LIANG ZHENG[1]

[1]*Triticeae Research Institute, Sichuan Agricultural University, Wenjiang, Chengdu, Sichuan 611130, People's Republic of China*
[2]*Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada, 960 Carling Ave, Ottawa, ON K1A 0C6, Canada*

[Qi P.-F., Chen Q., Ouellet T., Le C.-X., Ai J., Wei Y.-M., Wang J.-R. and Zheng Y.-L. 2013 Identification of 40k γ-secalin genes. *J. Genet.* **92**, 151–154]

Introduction

Secalins are the major seed storage proteins of rye (Gellrich *et al.* 2003), and can be classified into three major groups: the high-molecular-weight (HMW) secalins, γ-secalins and ω-secalins. The γ-secalins can be subdivided into 75k and 40k types on the basis of their molecular weight (Shewry *et al.* 1984). Secalins are usually encoded by gene families arranged in clusters. The genes encoding ω-secalins and 40k γ-secalins are located on the short arm of chromosome 1R, and those for the HMW secalins are located on the long arm of the same chromosome. The genes for 75k γ-secalins are located on the short arm of chromosome 2R in *Secale cereale* (Shewry *et al.* 1984).

Quantification of secalins indicated that the 75k γ-secalins contribute nearly half of the total seed storage protein, followed by the 40k γ-secalins (Gellrich *et al.* 2003). The primary structures of HMW-secalins, 75k γ-secalins and ω-secalins have been reported (Murray *et al*. 2001, De Bustos and Jouve 2003, Jiang *et al*. 2010). In contrast, the full nucleotide and amino acid sequences of the 40k γ-secalins are still unknown. Studies of amino acid composition, partial amino acid sequences, and molecular mass indicated that 40k γ-secalins are closely related to γ-gliadins of wheat (Gellrich *et al.* 2003, 2005). Here, we report for the first time the full sequences of 40k γ-secalins, and characterize their structures.

Materials and methods

Plant materials

The rye accessions PI240676 (*S. cereale* ssp. *cereale*), PI618662 (*S. cereale* ssp. *afghanicum*), PI573647 (*S. cereale* ssp. *tetraploidum*), PI561797 (*S. cereale*), PI240285 (*S. strictum* ssp. *anatolicum*), PI206992 (*S. strictum* ssp. *anatolicum*), PI618665 (*S. strictum* ssp. *ancestrale*), PI531829 (*S. strictum*), PI573648 (*S. vavilovii*) and PI618681 (*S. vavilovii*) were used for this study.

DNA extraction and gene cloning

Genomic DNA was extracted from fresh leaves of single fresh plant with a CTAB protocol (Doyle and Doyle 1987).

The PCR primers of Qi *et al.* (2009) were first used to isolate 40k γ-secalin genes from genomic DNA. The primers are as follows:

forward1: 5′-TATTAGTTAACGCAAATCCACC/TATG-3′;
forward2: 5′-CTTCACACAACTAGAGCACAAG-3′;
reverse1: 5′-GATGAATCAGCTAAGCAACGATG-3′;
reverse2: 5′-TCGTTACATCTATTGGTGCATCAG-3′

PCR amplification was conducted in a 25-μL volume, consisting of 100 ng genomic DNA, 100 μM of each dNTP, 1.5 mM Mg^{2+}, 2 pmol of each of the four primers, 0.75 U *Taq* polymerase with high fidelity (TianGen, Beijing, China) and 2.5 μL 10× buffer. The reactions were run in a PTC-240 (MJ Research, Ramsey, USA) thermal cycler with the following program: an initial step of 94°C for 4 min; 35 cycles of 94°C for 45 s, 57°C for 1 min and 72°C for 80 s; then a final step of 8 min at 72°C.

For the DNA samples for which PCR amplification failed, a new pair of primers (forward2: 5′-CTTCACA CAACTAGAGCACAAG-3′ and reverse3: 5′-CAATGAC CTGATCGCCTCCA-3′) was used under the same PCR conditions, except that 4 pmol of each of the two primers was used. This new primer pair missed the 3′ part of γ-secalin genes.

*For correspondence. E-mail: pengfeiqi@hotmail.com.

Keywords. rye; seed storage protein; structure.

The amplified products were separated on 1.5% agarose gel. The bands were recovered and the DNA was cloned into pMD-19T vector (Takara, Dalian, China). The recombinant vector was used to transform competent *Escherichia coli* JM109 cells. Colonies were detected by PCR using the same primers as for the initial amplification. Positive colonies were sequenced by a commercial provider (Invitrogen, Shanghai, China).

Sequence analysis

Sequences were aligned in DNAman v. 5.2.2 (Lynnon Biosoft, Pointe-Claire, Canada) using default parameters. The alignments were further improved by visual examination and manual adjustments. Neighbour-joining trees with bootstrap test (1000 replicates) were constructed for deduced amino acid sequences on the basis of Poisson correction, pairwise deletion of gaps using MEGA v. 5 (Tamura *et al.* 2011).

Results and discussion

Gene cloning

To determine the full gene sequence for 40k γ-secalins, we used the primers, which can successfully clone γ-gliadin genes from *Triticum* and *Aegilops* species, to clone γ-secalin genes from rye. Consequently, 16, 15 and five γ-secalin genes were isolated from *S. cereale*, *S. strictum* and *S. vavilovii* respectively (table 1). These gene sequences show a very close structural relationship to the γ-gliadin genes. Two of the 36 sequences (i.e. HQ875894 and HQ875932) have complete open reading frame (ORF). The remaining 34 have partial ORF, due to absence of the most-C-terminal domain in the fragments cloned. Eleven of the 36 sequences are putatively functional, and the other 25 are pseudogenes due to the presence of in-frame stop codons and frameshift mutations.

Identification of 40k γ-secalin genes

All the cloned γ-secalin genes are intronless. The estimated molecular masses for 40k and 75k γ-secalins are about 32 kDa and 52 kDa, respectively (Gellrich *et al.* 2003). Molecular masses of deduced mature proteins (minus signal peptide) of HQ875894 and HQ875932 are 21.3 kDa and 36.0 kDa (removing one in-frame stop codon), respectively, which are close to the 32 kDa of 40k γ-secalins, and much smaller than the 52 kDa of 75k γ-secalins. The higher molecular mass of 75k γ-secalins than that of γ-gliadins is a result of the addition of a repetitive sequence (figure 1). Comparison of amino acid sequences further suggested that the 36 sequences are 40k γ-secalin genes, since their deduced amino acid sequences have high homology with the reported partial amino acid sequences of 40k γ-secalins (Gellrich *et al.* 2005). Interestingly, there are three 75k γ-secalin genes, i.e. HQ266675, HQ266676 and HQ266677, with a much shorter repetitive domain than other 75k γ-secalins. Molecular masses of their deduced mature proteins are in the range 20.2–25.4 kDa, close to that of 40k γ-secalins.

The deduced amino acid sequences of putatively functional 40k and 75k γ-secalin genes were used to perform phylogenetic analysis. As a result, 40k and 75k γ-secalins

Table 1. The γ-secalin genes used in this study.

Species	Accession	Putatively functional		Pseudogene	Reference
40k γ-secalin					
S. cereale ssp. *cereale*	PI 240676	**HQ875966**		**HQ876000-HQ876002**	This study
S. cereale ssp. *afghanicum*	PI 618662	HQ875894	**HQ875957**	**HQ875979-HQ875981**	This study
S. cereale ssp. *tetraploidum*	PI 573647	**HQ875961**	**HQ875962**	**HQ875986-HQ875989**	This study
S. cereale	PI 561797			HQ875932	This study
S. strictum ssp. *anatolicum*	PI 240285	**HQ875958-HQ875960**			This study
S. strictum ssp. *anatolicum*	PI 206992	**HQ875963**	**HQ875964**	**HQ875995**	This study
S. strictum ssp. *ancestrale*	PI 618665			**HQ875990-HQ875994**	This study
S. strictum	PI 531829			**HQ875982-HQ875985**	This study
S. vavilovii	PI 573648	**HQ875965**		**HQ875996-HQ875998**	This study
S. vavilovii	PI 618681			**HQ875999**	This study
75k γ-secalin					
S. cereale	Unknown	AF20184			Murray *et al.* (2001)
S. cereale	Unknown	**X02602**			Kreis *et al.* (1985)
S. cereale	Unknown	EU368041			Unpublished
S. cereale	PI 412949	EF432547			Chen *et al.* (2008)
S. strictum	PI 531829	EF432549			Chen *et al.* (2008)
S. strictum	Qinling	HQ266670-HQ266677		HQ266723-HQ266726	Unpublished
S. sylvestre	NGB 5073	EF432546			Chen *et al.* (2008)
S. vavilovii	PI 618680	EF432548			Chen *et al.* (2008)

Genes without full ORF are indicated in boldface.

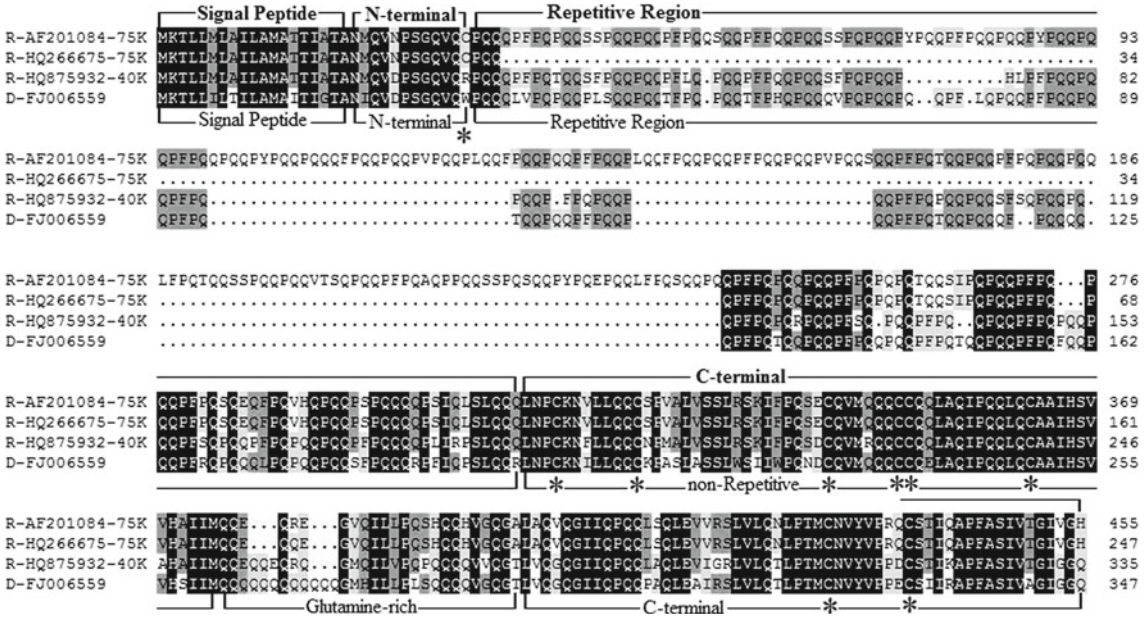

Figure 1. Alignment of γ-gliadin, 40k and 75k γ-secalin peptides. Asterisks at the bottom of the alignment indicate the positions of cysteine. AF201084 and HQ266675 are 75k γ-secalin genes; HQ875932 is a 40k γ-secalin gene; FJ006559 is a γ-gliadin gene from *Aegilops tauschii*.

were clearly separated into two major branches (figure 2), which reflects the significant difference in their primary structure.

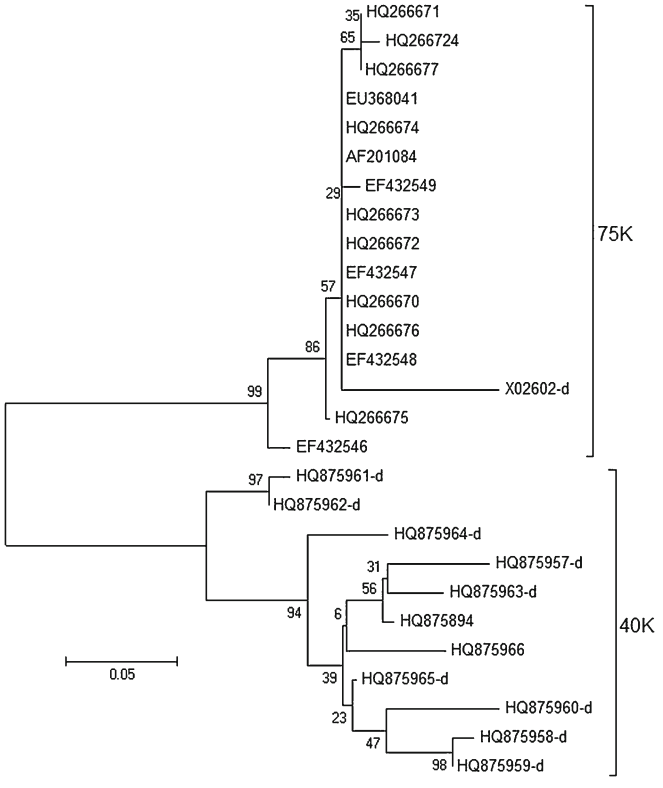

Figure 2. Phylogenetic analysis of the deduced peptides of γ-secalin genes. Accession numbers ending with the letter 'd' indicate genes with incomplete open reading frame.

Molecular mass is not a good discrimination factor for γ-secalins; and it is inappropriate to use the term 40k or 75k γ-secalins to describe some γ-secalins like HQ266675, HQ266676 and HQ266677, because of their special combination of a smaller molecular mass like the 40k γ-secalins and a primary structure like that of the 75k γ-secalins. Therefore, we recommend the use of 40k-type and 75k-type to classify γ-secalins according to their structural characteristics. Comparison of the nucleotide and amino acid sequences revealed very close structural relationship between 40k γ-secalins and γ-gliadins (figure 1), which confirms their classification in a common protein group.

Structural characteristics

The γ-gliadins and 40k γ-secalins share a signal peptide, an N-terminal domain, a highly variable central repetitive domain, a nonrepetitive domain containing six of the cysteine residues, a glutamine-rich region, and a C-terminal nonrepetitive domain containing the final two conserved cysteine residues (figure 1) (Anderson *et al.* 2001). The 75k γ-secalins contain a signal peptide, an N-terminal domain with a cysteine residue, followed by a central repetitive domain and a C-terminal domain with eight conserved cysteine residues (figure 1).

The repetitive domains of 40k and 75k γ-secalins are variable in length, and consist of a repeat unit. The reported typical unit for both γ-gliadins and 75k γ-secalins is $PFPQ_{1\text{-}2}(PQQ)_{1\text{-}2}$ (Anderson *et al.* 2001; Chen *et al.* 2008). However, residue substitutions, insertions/deletions within the repeat unit led to divergence of their repetitive domains (figure 1). A comparison of the glutamine-rich region of 40k

γ-secalins with the corresponding region in 75k γ-secalins (figure 1) suggests that the poly-Q within the glutamine-rich regions could be the result of a self-expansion of Q.

Most γ-gliadins contain eight cysteine residues that form four intramolecular disulphide bonds (Qi *et al.* 2009). The 40k γ-secalins identified here have eight cysteine residues, and show a conserved cysteine pattern similar to most γ-gliadins. Therefore, the 40k γ-secalins have four intramolecular disulphide bonds, as do γ-gliadins, consistent with the result that the major portion of 40k γ-secalins occurs in rye flour as monomers (Gellrich *et al.* 2003).

Taken together, the differences between 40k and 75k γ-secalins were found in the variation of poly-Q within the glutamine-rich region, in the modifications of the repetitive domains, and in the lack of a cysteine residue at the N-terminal domain.

Acknowledgements

This research was supported by the National Basic Research Program of China (973 Program 2011CB100100) and the National High Technology Research and Development Program of China (863 program 2011AA100103-02).

References

Anderson O. D., Hisa C. C. and Torres V. 2001 The wheat γ-gliadin genes: characterization of ten new sequences and further understanding of γ-gliadin gene family structure. *Theor. Appl. Genet.* **103**, 323–330.

Chen Q. J., Yuan Z. W., Zhang L. Q., Yan Z. H., Xiang, Z. G., Wan Y. F. *et al.* 2008 Molecular characterization and comparative analysis of four new genes from *Sec2* locus encoding 75 K γ-secalins of rye species. *J. Cereal Sci.* **48**, 111–116.

De Bustos A. and Jouve N. 2003 Characterization and analysis of new HMW-glutenin alleles encoded by the Glu-R1 locus of Secale cereale. *Theor. Appl. Genet*. **107**, 74–83.

Doyle J. J. and Doyle J. L. 1987 A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochem. Bull.* **19**, 11–15.

Gellrich C., Schieberle P. and Wieser H. 2003 Biochemical characterization and quantification of the storage protein (secalin) types in rye flour. *Cereal Chem.* **80**, 102–109.

Gellrich C., Schieberle P. and Wieser H. 2005 Studies of partial amino acid sequences of γ-40 K secalins of rye. *Cereal Chem.* **82**, 541–545.

Jiang Q. T., Wei Y. M., Andre L., Lu Z. X., Pu Z. E., Peng Y. Y. and Zheng Y. L. 2010 Characterization of omega secalin genes from rye, triticale and a wheat 1BL/1RS translocation line. *J. Appl. Genet*. **51**, 403–411.

Kreis M., Forde B. G., Rahman S., Miflin B. J. and Shewry P. R. 1985 Molecular evolution of the seed storage proteins of barley, rye and wheat. *J. Mol. Biol.* **183**, 499–502.

Murray F. R., Skerritt J. H. and Appels R. 2001 A gene from the *Sec2* (*Gli-R2*) locus of a wheat 2RS.2BL chromosomal translocation line. *Theor. Appl. Genet.* **102**, 431–439.

Qi P. F., Wei Y. M., Ouellet T., Chen Q., Tan X. and Zheng Y. L. 2009 The γ-gliadin multigene family in common wheat (*Triticum aestivum*) and its closely related species. *BMC Genomics* **10**, 168.

Shewry P. R., Bradberry D., Franklin J. and White R. P. 1984 The chromosomal locations linkage relationships of the structural genes for the prolamin storage proteins (secalins) of rye. *Theor. Appl. Genet.* **69**, 36–69.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S. 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

Received 19 June 2012; accepted 22 January 2013
Published on the Web: 10 April 2013