GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Isolation and phylogenetic analysis of novel γ-gliadin genes in genus *Dasypyrum*

G.R. Li[1]*, C. Liu[1]*, E.N. Yang[2] and Z.J. Yang[1]

[1]School of Life Science and Technology,
University of Electronic Science and Technology of China, Chengdu, China
[2]Crop Research Institute, Sichuan Academy of Agricultural Science,
Chengdu, China

*These authors contributed equally to this study.
Corresponding author: Z.J. Yang
E-mail: yangzujun@uestc.edu.cn

Genet. Mol. Res. 12 (1): 783-790 (2013)
Received June 20, 2012
Accepted November 31, 2012
Published March 13, 2013
DOI http://dx.doi.org/10.4238/2013.March.13.7

ABSTRACT. As the most ancient member of the wheat gluten family, the γ-gliadin genes are suitable for phylogenetic analysis among wheat and related species. Species in the grass genus *Dasypyrum* have been widely used for wheat cross breeding. However, the genomic relationships among *Dasypyrum* species have been little studied. We isolated 22 novel γ-gliadin gene sequences, among which 10 are putatively functional. The open reading frame lengths of these sequences range from 642 to 933 bp, and these putative proteins consist of five domains. Phylogenetic analyses showed that all *Dasypyrum* γ-gliadin gene sequences clustered in a large group; *D. villosum* and tetraploid *D. breviaristatum* γ-gliadin gene sequences clustered in a subgroup, while diploid *D. breviaristatum* γ-gliadin gene sequences clustered at the edge of the subgroup. All of the *Dasypyrum* γ-gliadin gene sequences were absent in three major T cell-stimulatory epitopes binding to HLA-DQ2/8 in celiac disease patients. Based on the phylogenetic analyses, we suggest that *D. villosum* and tetraploid *D. breviaristatum* evolved in parallel from a diploid ancestor *D. breviaristatum.*

Key words: *Dasypyrum*; γ-gliadin genes; Phylogenetic analysis

INTRODUCTION

The most abundant components of the seed storage proteins in wheat are gliadins and glutenins (Bartels and Thompson, 1983). The glutenin is mainly composed of high-molecular-weight (HMW) and low-molecular-weight (LMW) glutenin subunits, of which, HMW are encoded by genes from the long arms of homologous group 1 chromosomes (Payne, 1987), while LMW are encoded by genes from the short arms of homologous group 1 chromosomes (Singh and Shepherd, 1988). The components of gliadins are assigned to the α-, β-, γ-, and ω-subfamilies based on their electrophoretic patterns, DNA or protein structure and chromosomal locations (Masoudi-Nejad et al., 2002). The α- and β-subfamilies are clustered on the short arms of homologous group 6 chromosomes, while γ- and ω-subfamilies are located on the short arms of homologous group 1 chromosomes (Dubcovsky et al., 1997). The γ-gliadins have been considered to be the most ancient members of the wheat glutenin (Anderson et al., 2001). The molecular cloning and sequencing of these genes are helpful for better understanding the relationships of their structures and functions, and will provide vital information on gene evolution (Guo et al., 2010). In the Triticeae, γ-gliadin genes have been cloned from the genus *Triticum* (Anderson et al., 2001; Qi et al., 2009; Altenbach et al., 2010), *Aegilops* (Qi et al., 2009; Huang et al., 2010), *Thinopyrum* (Chen et al., 2009), and *Crithopsis* (Guo et al., 2010). However, no γ-gliadin genes have been cloned from the genus *Dasypyrum*.

The genus *Dasypyrum* consists of *D. villosum* (L.) Candargy (syn. *Haynaldia villosa* (L.) Schur) and *D. breviaristatum* (Lindb. F.) Frederiksen (syn. *H. hordeaceae* Coss. et Dur.). *D. villosum* is an annual allogamous diploid species (2n = 2x = 14, VV) mainly distributed from the Mediterranean to the Caspian Sea. *D. breviaristatum* consists of diploid (2n = 2x = 14, V^bV^b) and tetraploid (2n = 4x = 14, $V^bV^bV^bV^b$) cytotypes. They are perennial allogamous species restricted to two mountainous regions in northwest Africa and in Greece and Morocco (Gradzielewska, 2006a). Species of the genus *Dasypyrum* contain many agronomically useful genes, which could be of value in wheat breeding for multi-disease resistance, better quality and high yield (Gradzielewska, 2006b). However, the controversial relationships of *Dasypyrum* species have befogged the utilization of these species, especially *D. breviaristatum* (Yang et al., 2006; Liu et al., 2006, 2010). The prolamin genes have proven to be very useful in studying Triticeae evolution (Yan et al., 2006; Li et al., 2009). Therefore, in this research, we cloned the γ-gliadin genes from the species of the genus *Dasypyrum* to investigate their phylogenic relationships.

MATERIAL AND METHODS

Plant materials

Tetraploid *D. breviaristatum* (PI 516547) was obtained from Dr. Harold Bockelman, National Plant Germplasm System, USDA-ARS, Aberdeen, ID, USA. Diploid *D. breviaristatum* (99008-8) was provided by Dr. Shoji Ohta, Department of Bioscience, Fukui Prefectural University, Matsuoka, Yoshida, Fukui, Japan. *D. villosum* (TA10220) was obtained from Dr. W. Jon Raupp, Wheat Genetics Resource Center and Department of Plant Pathology, Throckmorton Plant Sciences Center, Kansas State University, Manhattan, KS, USA.

Primer design, PCR cloning and sequencing

Total genomic DNA was prepared from young leaves using the SDS protocol (Yang et al., 2005). The DNA concentration was determined using a Sizhumen DNA-protein photometer, and also by comparison with a known lambda DNA standard on an agarose gel. Primer synthesis and PCR protocol followed that of Chen et al. (2009). The target genes identified by PCR were excised from 1.0% agarose gels and purified using a gel extraction kit (Qiagen, Valencia, CA, USA). The purified products were ligated into the pT7 Blue R-Vector using T4 ligase, and then introduced into *Escherichia coli* DH5α by heat-shock transformation. Nucleotide sequencing was performed on a polyacrylamide gel with the ABI prism 377 sequencer (Perkin Elmer) as an automated fluorescent sequencing system.

Phylogenetic analyses

The γ-gliadin gene sequences, *Lophopyrum elongatum* (FJ040760 and FJ040757), *Triticum aestivum* (FJ006589 and FJ006611), *Aegilops searsii* (FJ006690, FJ006688 and FJ006689), *A. bicomis* (FJ006711 and FJ006712), and *A. speltoides* (FJ006697, FJ006701, FJ006699, FJ006694, FJ006670, and FJ006672), used as comparison controls were obtained from NCBI website (http://www.ncbi.nlm.nih.gov/). The *Dasypyrum* γ-gliadin sequences (JF441247-JF441267) cloned here were analyzed by the ORF finder program at the NCBI network service (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Sequences were aligned using the BioEdit software. All DNA sequences were aligned using CLUSTAL W version 1.8 (Thompson et al., 1994). Multiple alignment parameters were scored up to 12 for gap opening penalty and 0.1 for gap extension penalty. Alignments were confirmed manually using sequential pairwise comparisons. MEGA4 was used for calculating pairwise sequence divergences and nucleotide compositions, and for performing neighbor-joining analyses. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages and was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. A consensus tree was generated using 1000 bootstrap replicates (Tamura et al., 2007).

RESULTS AND DISCUSSION

Isolation and sequencing of *Dasypyrum* γ-gliadin sequences

The primer pair used was designed according to Chen et al. (2009) to amplify γ-gliadin sequences from genus *Dasypyrum* species. A total of 20 clones from each template were sequenced, resulting in 22 usable sequences (JF441246-JF441267), 8 from tetraploid *D. breviaristatum*, 8 from diploid *D. breviaristatum* and 6 from *D. villosum* (Table 1). The lengths of these sequences ranged from 642 to 933 bp. These sequences were highly homologous (85-99%) to the γ-gliadin genes deposited in the NCBI database library (Table 1), confirming that they were γ-gliadin sequences. Among these 22 sequences, 10 represented a full-open reading frame (ORF), and the remaining 12 sequences were classified as probable pseudogenes (Table 1).

Table 1. Peptide sequence analysis of the γ-gliadin sequences recovered in genus *Dasypyrum*.

Species, accession and genome formula	Sequences cloned (GenBank No.)	Fragment length (bp)	Deduced amino acid length	Repeat No. in domain II	No. of cysteine residues
D. breviaristatum (4x)	JF441246	867	289	14	8
PI 516547	JF441248	918	306	15	8
	JF441250	888	296	15	8
	JF441252	867	289	15	8
	JF441247*	825	-	-	-
	JF441249*	918	-	-	-
	JF441251*	825	-	-	-
	JF441253*	867	-	-	-
D. breviaristatum (2x)	JF441255	828	276	20	8
99008-8	JF441256	933	311	22	8
	JF441254*	909	-	-	-
	JF441257*	910	-	-	-
	JF441258*	909	-	-	-
	JF441259*	825	-	-	-
	JF441260*	913	-	-	-
	JF441261*	909	-	-	-
D. villosum	JF441263	867	289	15	8
TA10220	JF441264	867	289	15	8
	JF441265	900	300	16	8
	JF441266	642	214	12	8
	JF441262*	831	-	-	-
	JF441267*	899	-	-	-

(*) = pseudogene; (-) = no related information.

Deduced amino acid sequence analysis

Dasypyrum γ-gliadin genes cloned in the present study encode a polypeptide chain with lengths ranging from 214 to 311 amino acids. Deduced amino acid sequence alignment showed that the ORFs had highly similar structures (Figure 1). According to the sequence structure characteristics (Chen et al., 2009; Qi et al., 2009; Huang et al., 2010), the deduced amino acid sequences encoded by γ-gliadin genes are divided into domains I to V. Domain I has a conserved 12-residue stretch in the N-terminal region. Domain II includes 12-22 tandem repeats. Domain III contains 6/7 cysteine residues without a repetitive stretch. Domain IV is rich in glutamine residues. Domain V, a non-repetitive region, contains the last two conserved cysteines in the C-terminal region. All peptide chains encoded by *Dasypyrum* γ-gliadin genes had 8 cysteine residues and a 12 to 22 direct repeat in domain II (Table 1). The variable length parts of *Dasypyrum* γ-gliadin genes were domains II and IV, while domains I, III and V were conserved. Compared to Triteceae control sequences, *Dasypyrum* peptide chains had an additional QQQ(H)VGQGT sequence in domain IV, suggesting that QQQ(H)VGQGT, corresponding to CAGCAACAG(T)GTGGGTCAAGGTACT, was *Dasypyrum*-specific. Therefore, we could design specific molecular primers based on this part of the nucleotide sequence to obtain *Dasypyrum*-specific γ-gliadin gene markers for screening wheat-*Dasypyrum* cross offspring in wheat-breeding programs.

Phylogenetic analysis of the *Dasypyrum* γ-gliadin genes

For both plant breeding and evolutionary studies, tracing the origin of species and determining their genomic relationships with close ones provide vital information to guide

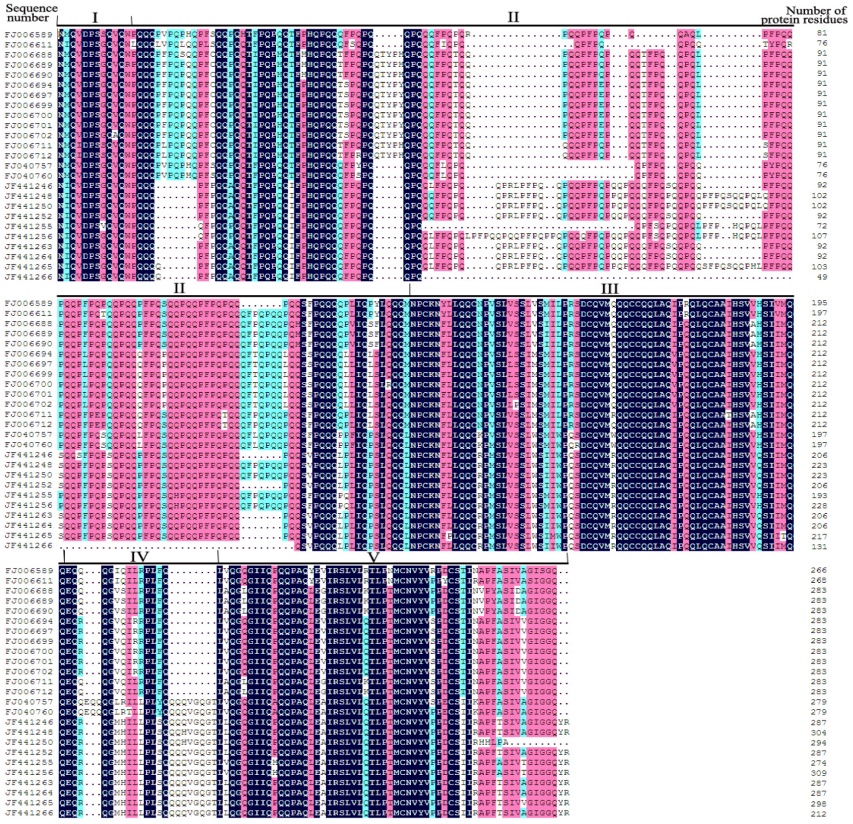

Figure 1. Comparison of the deduced amino acid sequences of the *Dasypyrum* γ-gliadin genes (the first two signal peptides are not included) with Triticeae-related control sequences. I to V represent the conserved N-terminal region, tandem repeat domain, non-repetitive domain, glutamine-rich domain, and C-terminal region, respectively.

the selection of desirable traits and to better understand how genomes function and evolve (Law, 1981). In the last several decades, research has focused on the relationship between *D. breviarisatatum* and *D. villosum*. According to Love (1984), the genomic formula of tetraploid *D. breviarisatatum* would be VVVV, which is the same as that of *D. villosum*. Similarly, this view was supported by biochemical evidence, such as the phenotypes of glutamic-oxaloacetic-transaminase, superoxide dismutase, alcohol dehydrogenase and esterase isozyme systems (Blanco et al., 1996). However, cytogenetic evidence such as fluorescence *in situ* hybridization patterns (Galasso et al., 1997; Uslu et al., 1999) and Giemsa C-banded karyotype (Frederiksen, 1991) seemed to be opposed to this view. Therefore, *D. breviarisatatum*'s origin is still ambiguous. Ohta et al. (2002) rediscovered diploid *D. breviarisatatum* in Morocco, providing an opportunity to re-estimate the evolutionary relationships of the *Dasypyrum* species. Based on the morphological and cytogenetic data, they concluded that the diploid cytotype is the most probable candidate for the ancestral form of the tetraploid cytotype (Ohta and Morishita, 2001; Ohta et al., 2002). In this research, phylogenetic analysis was performed on γ-gliadin amino acid sequences to construct a cladogram using Triticeae representative amino acid sequences as

controls. The repetitive regions of γ-gliadin consist of a long direct repeat and evolve rapidly, so they are not considered suitable for determining relatedness (Anderson et al., 2001). Therefore, the repetitive domains are not included here for constructing the cladogram. As shown in Figure 2, *T. aestivum*, *L. elongatum*, *A. searsii*, *A. bicomis*, and *A. speltoides* clustered in a subgroup according to their genome formula, implying that γ-gliadin genes may show species specificity. All *Dasypyrum* clustered in a big group, while *D. villosum* and tetraploid *D. breviaristatum* clustered in a subgroup, diploid *D. breviaristatum* clustered at the periphery of the subgroup (Figure 2), suggesting the following possibilities: 1) the relationships between *Dasypyrum* species are closer than for other Triticeae species; 2) *D. villosum* and tetraploid *D. breviaristatum* evolved in parallel, and 3) *D. villosum* and tetraploid *D. breviaristatum* might have originated from an ancestor diploid *D. breviaristatum*. The first two possibilities support our earlier report (Liu et al., 2010). The third view supports that of De Pace et al. (2011).

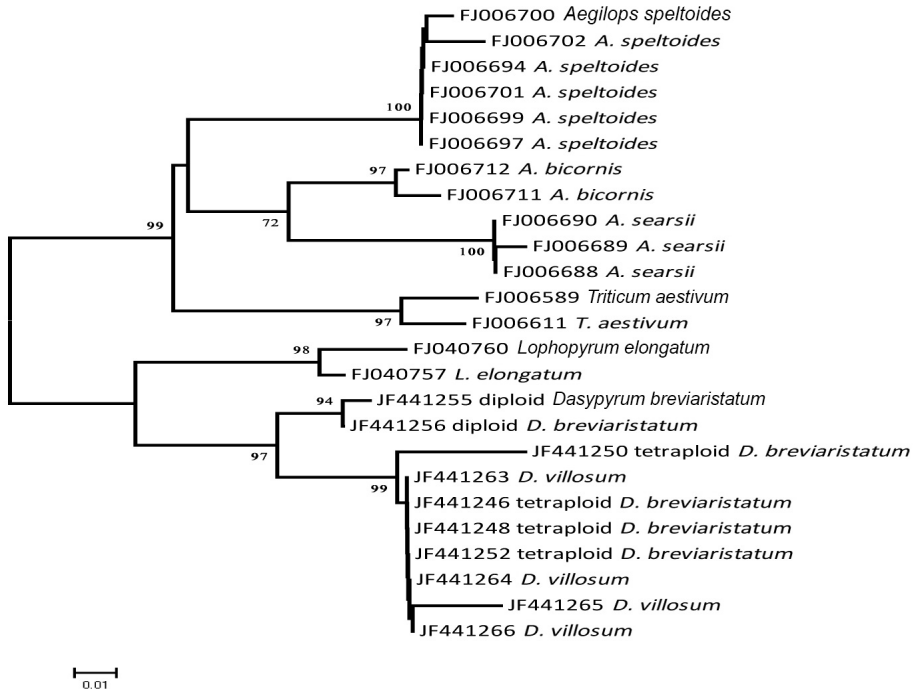

Figure 2. Evolutionary relationships based on the alignments of the amino acid sequences of 10 non-repetitive γ-gliadins from genus *Dasypyrum* and 15 related Triticeae controls by MEGA4 using the neighbor-joining method. Numbers above the branches show bootstrap frequencies based on 1000 bootstrap replicates.

Analysis of celiac disease (CD)-toxic epitopes

Gänzle et al. (2008) reported that three T cell-stimulatory epitopes for CD patients are 26mer (FLQPQQPFPQQPQQPYPQQPQQPFPQ), 16mer (LQPQQPFPQQPQQPYPQQPQ), and 13mer (FSQPQQQFPQPQ). Each epitope had its own position in the γ-gliadin protein. We search the perfect matches in the obtained full-ORF genes and in the pseudogenes to the

four epitopes of *Dasypyrum* γ-gliadin gene sequences. The single nucleotide polymorphism resulted in an amino acid change occurring in a particular epitope. All ORFs displayed the deletion of a glutamine (Q) in both type 2 and type 3 epitopes, the presence of arginine (R) instead of proline (P) in type 1 region, and the changes of several amino acids at position of type 1 and type 2 regions. Therefore, the results demonstrate that the set of epitopes are totally absent in the *Dasypyrum* γ-gliadin sequences.

Recently, van Herpen et al. (2006) reported that there are large differences in the content of predicted T cell epitopes (glia-α, glia-α2, glia-α9, glia-α20) in full-ORF genes and pseudogenes from the wheat α-gliadin gene sequences. Our previous study revealed that *Dasypyrum* α-gliadin sequences contain the glia-α epitope but lack the 3 other epitopes (Li et al., 2009). The present study revealed that the *Dasypyrum* γ-gliadin sequences lack the 3 major epitopes completely. Thus, the existence of large differences in the distribution of toxic gliadin genes provides opportunities to select and breed wheat varieties with less toxic epitopes and suitable for consumption by CD patients by introducing *Dasypyrum* chromatin to wheat breeding.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31171542) and Fundamental Research Funds for the Central Universities (#ZYGX2011J101, #ZYGX2011J095).

REFERENCES

Altenbach SB, Vensel WH and DuPont FM (2010). Analysis of expressed sequence tags from a single wheat cultivar facilitates interpretation of tandem mass spectrometry data and discrimination of gamma gliadin proteins that may play different functional roles in flour. *BMC Plant Biol.* 10: 7.

Anderson OD, Hsia CC 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.

Bartels D and Thompson RD (1983). The characterization of cDNA clones coding for wheat storage proteins. *Nucleic Acids Res.* 11: 2961-2977.

Blanco A, Simeone R, Resta P, Pace CD, et al. (1996). Genomic relationships between *Dasypyrum villosum* (L.) Candargy and *D. hordeaceum* (Cosson et Durieu) Candargy. *Genome* 39: 83-92.

Chen FG, Zhao F, Liu SW and Xia GM (2009). The γ-gliadin gene content of a derivative from a somatic hybrid between bread wheat and tall wheatgrass. *Mol. Breed.* 24: 117-126.

De Pace C, Vaccino P, Cionini PG, Pasquini M, et al. (2011). *Dasypyrum*. In: Wild Crop Relatives, Genomic and Breeding Resources, Cereals (Kole C, eds.). Chapter 4. Springer-Verlag, Heidelberg, 185-292.

Dubcovsky J, Echaide M, Giancola S, Rousset M, et al. (1997). Seed-storage-protein loci in RFLP maps of diploid, tetraploid, and hexaploid wheat. *Theor. Appl. Genet.* 95: 1169-1180.

Frederiksen S (1991). Taxonomic studies in *Dasypyrum* (Poaceae). *Nord. J. Bot.* 11: 135-142.

Galasso I, Blanco A, Katsiotis A, Pignone D, et al. (1997). Genomic organization and phylogenetic relationships in the genus *Dasypyrum* analysed by southern and *in situ* hybridization of total genomic and cloned DNA probes. *Chromosoma* 106: 53-61.

Gänzle MG, Loponen J and Gobbetti M (2008). Proteolysis in sourdough fermentations: mechanisms and potential for improved bread quality. *Trends Food Sci. Technol.* 19: 513-521.

Gradzielewska A (2006a). The genus *Dasypyrum*-part 1. The taxonomy and relationships within *Dasypyrum* and with *Triticeae* species. *Euphytica* 152: 429-440.

Gradzielewska A (2006b). The genus *Dasypyrum*-part 2. *Dasypyrum villosum*-a wild species used in wheat improvement. *Euphytica* 152: 441-454.

Guo ZF, Zhong M, Wei YM, Zhang L, et al. (2010). Characterization of two novel γ-gliadin genes encoded by K genome

of *Crithopsis delileana* and evolution analysis with those from Triticeae. *Genes Genomics* 32: 259-265.
Huang Z, Long H, Wei YM, Qi PF, et al. (2010). Characterization and classification of γ-gliadin multigene sequences from *Aegilops* section *Sitopsis*. *Cereal Res. Comm.* 38: 1-14.
Law CN (1981). Chromosome manipulation in wheat. *Chromosomes Today* 7: 194-205.
Li GR, Liu C, Zeng ZX, Jia JQ, et al. (2009). Identification of α-gliadin genes in *Dasypyrum* in relation to evolution and breeding. *Euphytica* 165: 155-163.
Liu C, Yang ZJ, Feng J, Zhou JP, et al. (2006). Systematic status of *Dasypyrum breviaristatum* in Triticeae based on RAPD analyses. *Triticeae Crop* 26: 11-15.
Liu C, Li GR, Sunish S, Jia JQ, et al. (2010). Genome relationships in the genus *Dasypyrum*: evidence from molecular phylogenetic analysis and *in situ* hybridization. *Plant Syst. Evol.* 288: 149-156.
Love A (1984). Conspectus of the Triticeae. *Feddes Rep.* 95: 425-521.
Masoudi-Nejad A, Nasuda S, Kawabe A and Endo TR (2002). Molecular cloning, sequencing, and chromosome mapping of a 1A-encoded ω-type prolamin sequence from wheat. *Genome* 45: 661-669.
Ohta S and Morishita M (2001). Genome relationships in the genus *Dasypyrum* (Gramineae). *Hereditas* 135: 101-110.
Ohta S, Koto M, Osada T, Matsuyama A, et al. (2002). Rediscovery of a diploid cytotype of *Dasypyrum breviaristatum* in Morocco. *Genet. Resour. Crop Evol.* 49: 305-312.
Payne PI (1987). Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Annu. Rev. Genet.* 38: 141-153.
Qi PF, Wei YM, Ouellet T, Chen Q, et al. (2009). The γ-gliadin multigene family in common wheat (*Triticum aestivum*) and its closely related species. *BMC Genomics* 10: 168.
Singh NK and Shepherd KW (1988). Linkage mapping of genes controlling endosperm storage proteins in wheat. 1. Genes on the short arms of group-1 chromosomes. *Theor. Appl. Genet.* 75: 628-641.
Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.
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.
Uslu E, Reader SM and Miller TE (1999). Characterization of *Dasypyrum villosum* (L.) Candargy chromosomes by fluorescent *in situ* hybridization. *Hereditas* 131: 129-134.
van Herpen TW, Goryunova SV, van der Schoot J, Mitreva M, et al. (2006). Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes. *BMC Genomics* 7: 1.
Yan ZH, Wei YM, Wang JR, Liu DC, et al. (2006). Characterization of two HMW glutenin subunit genes from *Taenitherum* Nevski. *Genetica* 127: 267-276.
Yang ZJ, Li GR, Feng J, Jiang HR, et al. (2005). Molecular cytogenetic characterization and disease resistance observation of wheat-*Dasypyrum breviaristatum* partial amphiploid and its derivatives. *Hereditas* 142: 80-85.
Yang ZJ, Liu C, Feng J, Li GR, et al. (2006). Studies on genome relationship and species-specific PCR marker for *Dasypyrum breviaristatum* in Triticeae. *Hereditas* 143: 47-54.