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

Identification and sequence analysis of grain softness protein in selected wheat, rye and triticale

M.A.S. Kharrazi and V. Bobojonov

Department of Genetics, Agriculture University of Tajikistan, Dushanbe, Tajikistan

Corresponding author: M.A.S. Kharrazi
E-mail: shirazi60@yahoo.com

Genet. Mol. Res. 11 (3): 2578-2584 (2012)
Received November 7, 2011
Accepted March 19, 2012
Published July 19, 2012
DOI http://dx.doi.org/10.4238/2012.July.19.1

ABSTRACT. Grain softness protein (GSP) is an important protein for overcoming milling and grain defenses in the innate immunity systems of cereals. The objective of this study was to evaluate and understand GSP sequences in selected wheat, rye and triticale. Using sequences for this gene from a sequence database, we performed clustering analysis to compare the sequences obtained from 3 germplasms with other studied sequences for GSP. The maximum difference between the Hirmand GSP genotype in wheat and the database sequences was 23% in EF109396 and EF109399. Most amino acid variation between the GSP sequences involved the same amino acids. The Nikita rye *GSP* gene showed 64% identity with DQ269918 and AY667063. The isoelectric point in the GSP of wheat and Lasko triticale was significantly higher than that of rye GSP. In addition, parameters such as optical density, grand average of hydrophobicity, percentage of hydrophobicity and hydrophilic amino acids, and number of alpha helices and beta sheets in GSP were similar in wheat and triticale but not in wheat and rye.

Key words: Grain softness protein; Wheat; Amino acid indexes; Rye; Triticale

INTRODUCTION

The physical hardness of grain is positively correlated with the milling characteristics, water absorption, and fermentation capacity of flours (Simeone et al., 2006). In the Pooideae subfamily, grain softness protein (GSP), puroindoline B (Ikai, 1980), and puroindoline A (Pin A) in the hardness locus influence the quality of grain (Edwards et al., 2010). Grain softness protein was first named friabilin because of its abundance in water-washed starch from friable soft wheat, its scarcity in hard wheat. The protein has been suggested to act as a "nonstick" surface between starch granules and the protein matrix (Gollan et al., 2007). The proteins produced by these genes are lipid-binding proteins. The *GSP-1* gene in hexaploid wheat is located in the A and B genomes in chromosome 5. The role of the GSP as a marker in grain texture in wheat and other cereals is still being debated (Giroux and Morris, 1998). Despite the name of this protein, some research has shown that the grain texture of wheat is unrelated to GSP (Tranquilli et al., 2002); *GSP-1* has potential antimicrobial activity because it has a high proportion of Arg-Lys-Trp-Phe as well as a high isoelectric point (pI) related to puroindoline, which implicates it in grain defense rather than texture quality (Phillips et al., 2010). Furthermore the presence of GSP in durum wheat shows that GSP does not have major role in grain softness in wheat and other cereals.

Comprehensive research is needed to assess the variation of GSP sequences in other cereals, including rye and triticale, to help elucidate whether important quality traits such as grain hardness are determined in those crops the way they are in wheat. Little information is available regarding the grain quality of rye as a second grain for bread production and alcohol distillates. Rye has a very soft endosperm phenotype with a high expression level of hardness genes (Morrison et al., 1992; Jolly et al., 1996). In triticale, despite resistance to hard conditions and many beneficial traits, the poor quality and quantity of the flour make it more useful as a forage crop than a product for human use. In this study GSP genes from wheat, rye, and triticale were amplified and characterized, and homology and sequence variations for this gene within and between the 3 crops were evaluated using clustering and bioinformatics tools.

MATERIAL AND METHODS

This study was carried out in the biotechnology research center of the University of Zabol, Iran, in November 2010. The Hirmand wheat landrace, the Nikita cultivar for rye, and the Lasko triticale variety were cultured in a greenhouse for 10 days. DNA was extracted from 0.2 g pooled leaf tissue using the procedure of Dellaporta et al. (1983) with some modifications. The quality and quantity of DNA was measured using 1% gel electrophoresis and spectrophotometry. The conditions for polymerase chain reaction (PCR) were identical for the 3 types of DNA and included 1 U Taq DNA polymerase, 10 pmol of each primer, 200 µL deoxyribonucleotide triphosphates, 3 mM $MgCl_2$, and 20, 25, and 50 ng genomic DNA from wheat, rye, and triticale, respectively, in a 25-µL final volume. PCR was carried out in a Mastercycler machine under conditions consisting of 5 min at 94°C and 35 cycles of 30 s at 94°C, 1 min at 49°-52°C, and 1 min at 72°C. A final extension of 8 min at 72°C was also performed. After amplification, the PCR product for each genome was individually subjected to 1% agarose gel electrophoresis, and specific bands were purified from the gel using a Gen Jet gel extraction kit (Fermentas) as specified in the manufacturer manual. Specific bands were sent to Farayand Danesh Arian Company for sequencing.

The results were analyzed using CLC Main Workbench version 5. The motif structures of studied sequences were analyzed in the Pfam database (http:/pfam.sanger.ac.uk/ [accessed 27 January 2012]). Sequences analyses of characteristics such as aliphatic index (Ikai, 1980), estimated half-life (Bachmair et al., 1986; Gonda et al., 1989), extinction coefficient (Gill and von Hippel, 1989), atomic composition, total number of negatively (Asp + Glu) and positively charge residues (Arg + Lys), amino acid distribution, and dipeptide distribution were performed using the nucleotide and protein analyses options of CLC Main Workbench. For gene sequence analyses, we created an alignment using very accurate options of 10 for the gap open cost and 1 for the gap extension cost with free for the gap cost (because we aligned differently sized PCR products) using options in CLC Main Workbench. Maximum link hood phylogeny was designed between and within the GSP gene sequences of wheat, rye, and triticale obtained in this study and those in the National Center for Biotechnology Information (NCBI) database with parameters including unweighted pair group method analysis with the substitution model results (Jukes and Cantor, 1969).

RESULTS AND DISCUSSION

Wheat GSP

Using specific primers for *GSP-1*, we extracted a 464-bp PCR product on agarose gel. GSP-b1 of the Hirmand wheat cultivar was aligned with 11 accessions from the NCBI database. The maximum similarity coefficient was 81.59% between *GSP-1* in these accessions and EF109429. This sequence was extracted from a study by Haudry et al. (2007), who demonstrated the existence of 2 groups of GSP based on the number of polymorphisms. The rate of GSP polymorphism in the wild type of *Triticum aestivum* is lower than that of the cultivars that have been selected for culture. The maximum distance for GSP-b1 was 23%, between Hamoon and the EF109396 and EF109399 accessions. No significant differences were found between these sequences, but a single nucleotide change can alter conformation and protein function. At position 9 in the nucleotide sequence, guanine is replaced by adenine in Hamoon, Ef109429, and EF109421 compared with the consensus sequence. A phylogenic tree was constructed to illustrate similarity and distance using unweighted pair group method analysis (Figure 1). GSP obtained from Hamoon created a separated sequence with significant dissimilarity from other sequences. This variety originated in the Sistan Province in southeastern Iran, and specific climate and selective breeding programs have caused this distance between Hamoon and the accession in the study by Haudry et al. (2007). In amino acid sequence alignment, we considered our sequence, with frame +3 for translation, and 16 accessions from the NCBI database. The first protein-level polymorphism was at position 15, where the polar amino acid threonine was replaced with the nonpolar alanine. We found 6 amino acid differences between our sequence and the consensus sequence.

The upper part of Table 1 shows homology, and the lower part shows distance between GSP sequences. Wilkinson et al. (2008) have reported that only 38-42% identity occurs between *GSP-1* and puroindolines A and B in the wheat variant. The variants also differ from *GSP-1* in the tryptophan loop region. These tryptophan residues are critical for the binding of GSP and puroindolines to lipids, which is the reason they affect grain texture. Isolation and molecular characterization of oat complementary DNA clones encoding proteins structurally

related to wheat GSP have been evaluated (Tanchak et al., 1998) and have shown that tryptophanins (the name used for GSP and puroindolines in the study) were present in oat, wheat, barley, and rye but not in developing rice seed.

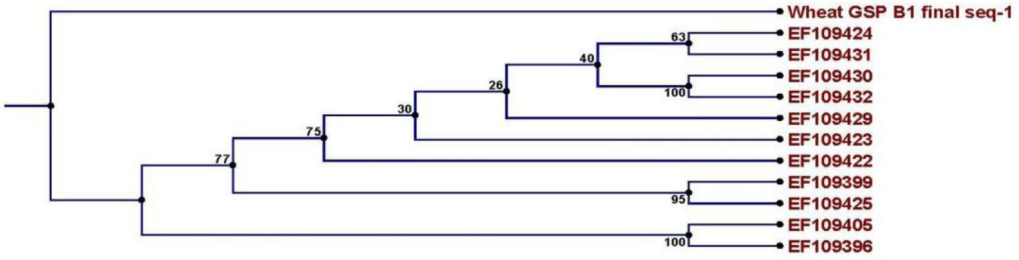

Figure 1. Dendrogram obtained for wheat grain softness protein (GSP) and 11 accessions from the NCBI database using UPGMA method.

Table 1. Specific primers to amplify grain softness protein in wheat, rye and triticale.

	Forward sequence	Reverse sequence	Annealing temperature (°C)	Self-annealing (°C)	GC (%)
Wheat	TTCCTCCTAGCTTTCCTT	GTGATCGGGATGTTGCAGT	59	16	55
Rye	CCTTCTTCCTCCTAGCTTT	TGCATCGTTCTGGCTTTT	55	10	47
Triticale	TTCCTCCTAGCTTTCCTT	GTGATCGGGATGTTGCAGT	59	16	55

Because these variations occur in the first part of the sequence under study, they can be created by error in the sequencing system. Also, the majority of these differences occur with the same amino acids-for instance, positions 114 in our target and the consensus sequence are valine and leucine, respectively. Both of these amino acids are nonpolar with aliphatic side chain. The pI for these sequences varied from pH 5.5 for AAX33220 to a pH 9 for CAA56598 but was close to 8 in Hirmand GSP. Maximum hydrophobicity of GSP occurred in amino acids in positions 25-33 and 34-54 (Table 2).

Table 2. Result of alignment between the Hamoon wheat grain softness protein (GSP) and 11 accession obtained from the NCBI (Haudry et al., 2007).

Column	1	2	3	4	5	6	7	8	9	10	11	Hamoon GSP
1		100	99.81	100	100	99.81	99.81	99.62	99.62	98.67	98.67	81.4
2	0		99.81	100	100	99.81	99.81	99.62	99.62	98.67	98.67	81.4
3	0	0		99.81	99.81	99.62	99.62	99.43	99.43	98.48	98.48	81.59
4	0	0	0		100	99.81	99.81	99.62	99.62	98.67	98.67	81.4
5	0	0	0	0		99.81	99.81	99.62	99.62	98.67	98.67	81.4
6	0	0	0	0	0		99.62	99.43	99.43	98.48	98.48	81.21
7	0	0	0	0	0	0		99.43	99.43	98.48	98.48	81.21
8	0	0	0.01	0	0	0.01	0.01		100	99.05	99.05	81.4
9	0	0	0.01	0	0	0.01	0.01	0		99.05	99.05	81.4
10	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.01	0.01		100	80.46
11	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0		80.46
Hamoon GSP	0.21	0.21	0.21	0.21	0.21	0.22	0.22	0.21	0.21	0.23	0.23	

GSP in rye

The Nikita cultivar was selected to study GSP in rye. As shown in Figure 2, high homology occurs between this sequence and 2 accessions obtained from the NCBI database and other studies (Beecher et al., 2002; Morris and Bhave, 2008). The complexity plot shows that maximum complexity occurs in positions 50-100 and 100-130. GC and AC were the maximum and minimum dinucleotides, respectively, in the GSP of Nikita rye. The GSP sequence chosen for study had 64% similarity with 2 accessions, DQ269918 and AY667063. The first difference occurred at position 20, where the rye GSP could be each of four nucleotides and the other 2 were adenine and guanine, respectively. Some studies have shown that the GSP in rye is a member of 25 cysteine-rich proteins (Lillemo et al., 2006; Massa and Morris, 2006). The pI for the amino acid sequence with frame +1 was 8.5. Our search of Pfam revealed no known patterns for GSP in rye, but one study has shown that rye GSP originates from 1 of 3 copies of wheat GSP (Simeone et al., 2006).

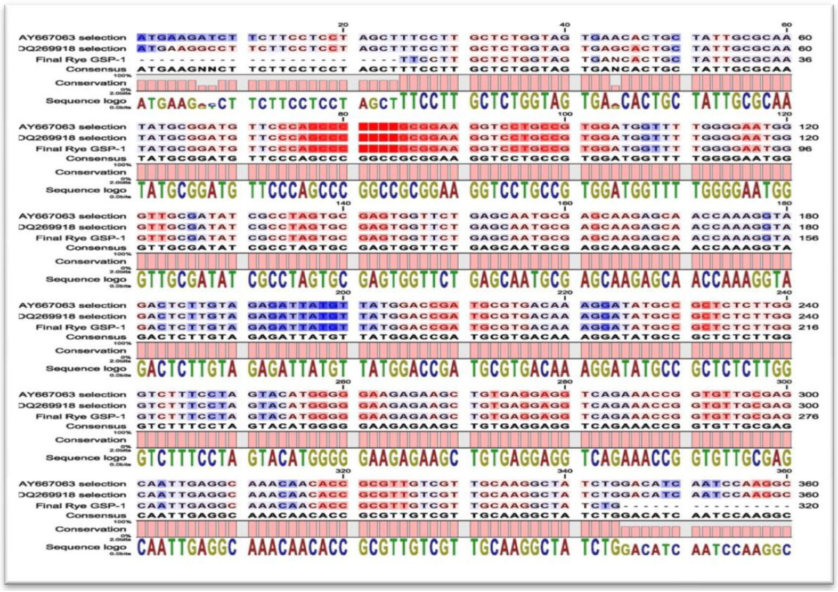

Figure 2. Result of alignment between DNA sequences in Nikita rye grain softness protein (Final Rye GSP-1) two accessions from database and consensus sequences.

Similarities among wheat, rye, and triticale

Few reports address the origin of GSP in triticale. One study has reported that it originated from a maternal source (wheat) (Massa and Morris, 2006). The main reason for the differences in pI between the GSP in wheat and those of rye and triticale is the high percentage of glutamic acid and aspartic acid in the GSPs of rye and triticale (Table 3). The thermo stability index in rye GSP was significantly higher than that in wheat and triticale GSPs because 18% of the total amino acids in rye are valine and alanine compared with 13 and 11% in wheat and

triticale, respectively. Extinction coefficient is an index of absorption at a specific wavelength in distinct proteins. This parameter is important for the purification and identification of a protein. Gill and von Hippel (1989) have shown that extinction coefficient can be calculated from the amino acid sequences of a protein. We measured this rate in relation to a molar extinction coefficient in tyrosine, tryptophan, and cysteine at specific wavelengths.

Table 3. Sequence specification of amino acid grain softness protein of wheat, triticale and rye.

Amino acid characteristics	Wheat	Rye	Triticale
Size of amplicon (bp)	143	106	143
Weight (g)	15.98	11.76	15.94
Isoelectric point (pH)	8.23	5.53	5.08
Aliphatic index	57.75	63.09	55.65
Extinction coefficient	34	20.11	34.11
Optical density	2.13	1.7	2.14
Atomic frequency HCNO	8-9-41-39	8-9-49-31	31-8-10-3
Hydrophobic frequency (%)	43	45	42
Hydrophilic frequency (%)	32	29	37
Negative-charge frequency (%)	11	13	11
Positive-charge frequency (%)	11	11	8
Maximum frequency of amino acid	Serine 9	Valine 9	Serine 9
Number of alpha helix	9	3	9
Number of beta sheet	2	3	2
Gravy	-0.49	-0.23	-0.49

For our purpose, the comparison of optical density was superior to that of other parameters because of the variation in the size of the GSP products. The optical densities for wheat and triticale were 2.13 and 2.4, respectively, and 1.7 in rye (see Table 3). This index is one of the most important indices showing that GSP in triticale originated from wheat rather than from rye. The same results were obtained in the grand average of hydrophobicity index (see Table 3). The percentages of hydrophobic amino acids, including alanine, phenyl alanine, glycine, isoleucine, leucine, methionine, proline, valine, and tryptophan in wheat, triticale, and rye for GSP were 39, 37, and 27, respectively. The frequency of hydrophilic amino acids occurs in the same pattern for GSP in wheat, rye, and triticale as does that of acidic amino acids, which make up 11% of wheat and rye GSP and 8% of rye GSP. Interestingly, the number of alpha helix and beta sheets was the same in wheat and triticale. All data in this study show that a strong positive correlation exists between GSP in triticale and wheat, and we conclude that the origin of GSP in triticale is the maternal source, wheat.

REFERENCES

Bachmair A, Finley D and Varshavsky A (1986). *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* 234: 179-186.

Beecher B, Bowman J, Martin JM, Bettge AD, et al. (2002). Hordoindolines are associated with a major endosperm-texture QTL in barley (*Hordeum vulgare*). *Genome* 45: 584-591.

Dellaporta S, Wood J and Hicks J (1983). A plant DNA minipreparation: version II. *Plant Mol. Biol. Rep.* 1: 19-21.

Edwards MA, Osborne BG and Henry RJ (2010). Puroindoline genotype, starch granule size distribution and milling quality of wheat. *J. Cereal Sci.* 52: 314-320.

Gill SC and von Hippel PH (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182: 319-326.

Giroux MJ and Morris CF (1998). Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline a and b. *Proc. Natl. Acad. Sci. U. S. A.* 95: 6262-6266.

Gollan P, Smith K and Bhave M (2007). Gsp-1 genes comprise a multigene family in wheat that exhibits a unique combination of sequence diversity yet conservation. *J. Cereal Sci.* 45: 184-198.

Gonda DK, Bachmair A, Wunning I, Tobias JW, et al. (1989). Universality and structure of the N-end rule. *J. Biol. Chem.* 264: 16700-16712.

Haudry A, Cenci A, Ravel C, Bataillon T, et al. (2007). Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol. Biol. Evol.* 24: 1506-1517.

Ikai A (1980). Thermostability and aliphatic index of globular proteins. *J. Biochem.* 88: 1895-1898.

Jolly CJ, Glenn GM and Rahman S (1996). GSP-1 genes are linked to the grain hardness locus (Ha) on wheat chromosome 5D. *Proc. Natl. Acad. Sci. U. S. A.* 93: 2408-2413.

Jukes T and Cantor C (1969). Evolution of Protein Molecules. Mammalian Protein Metabolism. Academic Press, New York.

Lillemo M, Chen F, Xia X, William M, et al. (2006). Puroindoline grain hardness alleles in CIMMYT bread wheat germplasm. *J. Cereal Sci.* 44: 86-92.

Massa AN and Morris CF (2006). Molecular evolution of the puroindoline-a, puroindoline-b, and grain softness protein-1 genes in the tribe Triticeae. *J. Mol. Evol.* 63: 526-536.

Morris CF and Bhave M (2008). Reconciliation of D-genome puroindoline allele designations with current DNA sequence data. *J. Cereal Sci.* 48: 277-287.

Morrison WR, Greenwell P, Law CN and Sulaiman BD (1992). Occurrence of friabilin, a low molecular weight protein associated with grain softness, on starch granules isolated from some wheats and related species. *J. Cereal Sci.* 15: 143-149.

Phillips RL, Palombo EA, Panozzo JF and Bhave M (2010). Puroindolines, Pin alleles, hordoindolines and grain softness proteins are sources of bactericidal and fungicidal peptides. *J. Cereal Sci.* 53: 112-117.

Simeone MC, Gedye KR, Mason-Gamer R, Gill BS, et al. (2006). Conserved regulatory elements identified from a comparative puroindoline gene sequence survey of *Triticum* and *Aegilops* diploid taxa. *J. Cereal Sci.* 44: 21-33.

Tanchak MA, Schernthaner JP, Giband M and Altosaar I (1998). Tryptophanins: isolation and molecular characterization of oat cDNA clones encoding proteins structurally related to puroindoline and wheat grain softness proteins. *Plant Sci.* 137: 173-184.

Tranquilli G, Heaton J, Chicaiza O and Dubcovsky J (2002). Substitutions and deletions of genes related to grain hardness in wheat and their effect on grain texture. *Crop Sci.* 42: 1812-1817.

Wilkinson M, Wan Y, Tosi P, Leverington M, et al. (2008). Identification and genetic mapping of variant forms of puroindoline b expressed in developing wheat grain. *J. Cereal Sci.* 48: 722-728.