J. Agr. Sci. Tech. (2020) Vol. 22(3): 863-874

Identification of Genome Specific Sequence Motifs in α-Gliadins and Wheat Accessions with Less Celiac Disease Epitopes

S. Singh[1*], S. Ram[2], and S. Narwal[2]

ABSTRACT

Among gliadins, α-gliadins are important active proteins in triggering celiac disease in human beings owing to the presence of toxic epitopes. A set of 177 *α-gliadin* gene sequences and the corresponding proteins were analyzed. Twenty accessions of hexaploids including 1, 14, and 5, respectively representing *A, B,* and *D,* with no intact CD-epitopes in *α*-gliadins, were identified. Twenty-two and 13 conserved motifs in non-repetitive domains NR1 and NR2, respectively, of *α*-gliadins differentiated all the amino acid sequences encoded by *A* genome of both diploids and hexaploids. Most of the amino acid sequences encoded by *D* genome (70 of 75 in hexaploids and 13 of 16 in diploids) could be identified by 22 amino acid motif. Large variations and lesser number of intact CD-epitopes was observed for *α*-gliadins belonging to *B* genome. As compared to diploids, repeat length of polyglutamine repetitive domain QII of *B* genome was lower in hexaploids indicating loss of Q residues during evolution of hexaploid wheat. The information can be used in assigning any *α*-gliadin sequences onto *A, B,* and *D* genomes and identifying wheat accessions with lesser CD-epitopes. The result presented here will be useful for the wheat improvement programs aiming for the management of celiac disease in human beings.

Keywords. Conserved motifs, Genome, *Triticum aestivum,* Wheat improvement.

INTRODUCTON

Wheat is one of the important cereal crops in the world providing energy and nutrition to human beings. Large numbers of end use products such as chapati, bread, biscuit, noodle, and pasta products are made from wheat because of the presence of gluten. Gluten is a visco-elastic complex formed when wheat flour is mixed with water. Gluten is formed by the interactions of gluten proteins in which millions of subunits of glutenins and gliadins are linked through di-sulphide linkages. The glutenins are subdivided into High Molecular Weight (HMW) and Low Molecular Weight (LMW) glutenins and the gliadins into *α, β, γ* and *ω* gliadins (Shewry and Halford, 2002). However, some components of gluten induce celiac disease (gluten intolerance) in susceptible human populations across the world. At present, many gluten derived T-cell stimulatory peptides are known which originate from the *α* and *γ* gliadins, and the HMW and LMW glutenins (Sjostrom *et al.*, 1998; Vande Wal *et al.*, 1998; Paulsen *et al.*, 1995; Molberg *et al.*, 2003; Anderson *et al.*, 2012). The *α*-gliadins, with average molecular weight of 31 kDa, are most abundant and represent 15–30% of the total wheat grain proteins. Therefore, *α*-gliadins are the most consumed storage proteins by human beings (Chen *et al.* 2008; Gu *et al.* 2004; Van Herpen *et al.*, 2006).

[1] Lovely Professional University, Phagwara-144411, Punjab, India.
[2]Indian Institute of Wheat & Barley Research (IIWBR), Karnal-132001 (Haryana), India.
*Corresponding author; e-mail: sanjaydbtster@gmail.com

The *α*-gliadins are reported to be the most active proteins in triggering celiac disease owing to the presence of several peptides, referred to as toxic epitopes, that constitute the main toxic components in celiac disease (Arentz-Hansen *et al.,* 2000; Vader *et al*., 2003; Vaccino *et al.,* 2009; Li *et al.,* 2010; Xie *et al.,* 2010; Kawaura *et al.,* 2012; Salentijn *et al.* 2013; Qi *et al.,* 2013; Li *et al.,* 2014a; Li *et al.,* 2014b). Since the only efficient therapy for celiac disease patients is a life-long gluten-free diet (Mc Manus and Kelleher, 2003), foods made from wheat flour with no or a few toxic epitopes are likely to be better tolerated, thereby improving a celiac disease patient's quality of life (Molberg *et al.* 2005; Spaenij-Dekking *et al.,* 2005). However, this remains a formidable challenge to develop "celiac-safe" wheat due to the complex multi-genic control of gluten protein composition (Shewry and Tatham, 2016).To begin with, there is need to identify wheat accessions with less number of celiac disease toxic epitopes (CD-epitopes). Therefore, knowledge of the number of CD-epitopes in *α*-gliadins and their location with respect to *A, B,* and *D* genomes is required. Though earlier reports by Van Herpen *et al.* (2006) and Vader *et al.* (2003) showed that size and distribution of toxic epitopes and the number of polyglutamine residues could be used to assign *α*-gliadin sequences to specific chromosomes, this could not differentiate *α*-gliadin sequences from large numbers of wheat accessions in this study. In this investigation, efforts were aimed to identify amino acid motifs differentiating *α*-gliadin sequences representing *A, B,* and *D* genomes. This would help to identify accessions with nil or fewer number of CD-epitopes representing diploid progenitors and hexaploid wheats.

MATERIALS AND METHODS

A set of 177 *α-gliadin* gene sequences and the corresponding proteins were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/protien/). The sequences represented112 hexaploid bread wheat (*T. Aestivum*) genotypes and 65 ancestral diploid progenitors of *A* (*Tmonococcum*, 30), *B* (*Aegilopsspeltoides& A. Searsii*, 19) and *D* (*Aegilopstauschii*, 16) genomes. The sequences were assembled and aligned using ClustalW (Thompson *et al.,* 1994) and manually curated using BioEdit version 7.0.9.0 (http://www.mbio.nesu.edu/BioEdit/Bioedit.html). Nucleotide diversity analysis of *α-gliadin* genes was conducted by Tajima *D* test using MEGA 6.0 (Tajima, 1989). The Multiple Sequence Alignment (MSA) of all the *α*-gliadin proteins was performed to construct a phylogenetic tree by MEGA (version 6.0) using default parameters and Maximum Likelihood (ML) method (Tamura *et al.,* 2013). The stability of branch nodes in the ML-tree was measured by performing bootstrap test of 1,000 replications to ensure a high confidence range and accuracy. The phylogenetic tree was further supported by Multiple Expectation-Maximization for Motif Elicitation (MEME) (Timothy *et al.,* 1994). About twenty different conserved motifs ranging from6 and 50 amino acids in length were detected by MEME software tool.

The identification of the four major immunogenic epitopes (Glia-α, α2, α9 and α20) and polyglutamine repeats (QI & QII) in *α*-gliadins was done as per the method of Van Herpen *et al.* (2006).The t-test was used for identifying the significant differences in the repeat length of polyglutamine (QI & QII) repeats in *α*-gliadins encoded by different genomes.

RESULTS

Identification and Frequency Distribution of Four Major Celiac Disease Toxic Epitopes in *α*-Gliadins

The entire set of *α*-gliadin protein sequences belongs to hexaploid bread wheat (*T. aestivum*) and their ancestral *A, B* and *D*

genome progenitors were retrieved and analyzed for the presence of CD-epitopes (Glia-α, α2, α9 and α20). Distribution of *α*-gliadin in hexaploid wheat genome was identified based on the three groups developed in the phylogenetic tree. The details are given in the section on phylogenetic analysis. The Glia-α (QGSFQPSQQ) was present in the second non- repetitive (NR2) domain, while Glia-α2 (PQPQLYPQ), Glia-α9 (PFPQPQLPY) and Glia-α20 (FRPQQPYPQ) were present in the first repetitive domain. There were large variations in the number of intact (T cell stimulatory) epitopes of *α*-gliadins representing *A*, *B* and *D* genomes of bread wheat and its progenitors (Table 1). Intact Glia-αencoded by *A* genome was absent in both the diploid and hexaploid sequences. The variable Glia-α was generated by a change of 5^{th} amino acid from *Q* (intact form) to *R* in all the Glia-α epitopes. Intact Glia-α2 encoded by *A* and *B* genomes was absent in all the diploid progenitor sequences. However, both the Glia-α and Glia-α2 encoded by *D* genome were present in higher frequency in *D* genome progenitors as well as in hexaploids. Although Intact Glia-α9 was absent in *B* genome progenitors, it was present in large numbers in *A* and *D* genome progenitors. In hexaploids, intact Glia-α9 was present more frequently in gliadins encoded by *D* genome. Intact Glia-α20 was present less frequently in *α*-gliadins encoded by *B* genomes of both diploids and hexaploids. Overall, the frequency of all four intact CD epitopes was higher in *α*-gliadins of both *D* genome progenitor *Aegilops tauschii* and encoded by *D* genome in hexaploids (Table 1).

Phylogenetic Analysis and Assignment of *α*-Gliadin Sequences onto by *A, B* and *D* Genomes

Phylogenetic tree of *α*-gliadins from 177 accessions was constructed by MEGA 6.0 to identify genome specific sequences (Figure 1). The tree exhibited three distinct groups of α-gliadins encoded by three genomes with a few exceptions. The MEME analysis also showed three distinct groups of *α*-gliadin sequences representing *A, B* and *D* genomes (Figure 2). Both the phylogenetic trees by MEGA 6.0 and the motif distribution by MEME exhibited similar pattern of groups and subgroups. Many motifs were found to be conserved and appeared in all the groups. These conserved motifs could be the essential elements determining the *α*-gliadin family's common molecular function. Of the 112 hexaploid sequences of *α*-gliadins, 11 represented *A* genome, 26 *B* genome and 75 *D* genome groups. The presence of three distinct groups is consistent with the hypothesis that *α-gliadin* gene family expansion occurred after the ancestors separated into the three *Triticum* genomes.

All the protein sequences were further

Table 1. Genome-wise distribution of T cell stimulatory (Intact) toxic epitopes presents in hexaploid bread wheat and their diploid genome progenitors. Genome-wise distribution in hexaploid sequences was identified based on the three groups developed in the phylogenetic tree. The details are given in the section on phylogenetic analysis.

Species	Genome	Glia-α	Glia-α2	Glia-α9	Glia-α20	N[a]
Diploid Progenitors						
T. monococcum	A	0	0	30	24	30
A. speltoides + *A. searsii*	B	16	0	0	2	19
A. tauschii	D	14	12	16	15	16
Hexaploid Wheat						
T. aestivum	A	0	1	6	6	11
T. aestivum	B	4	7	7	3	26
T. aestivum	D	60	60	58	42	75

[a] N is the total Number of sequences used in the analysis.

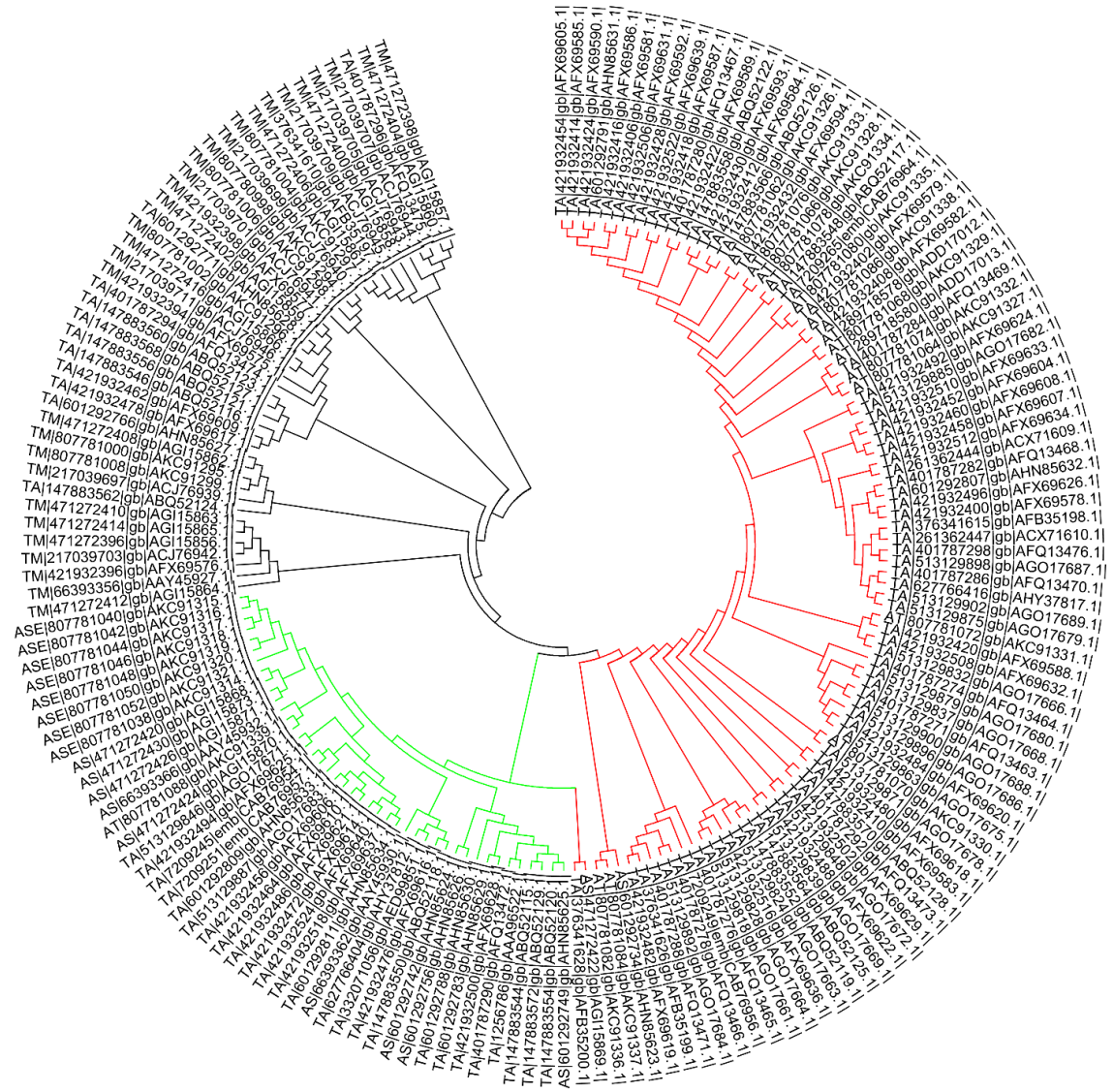

Figure 1. Phylogenetic analysis of amino acid sequences of α-gliadins representing 177 accessions of diploid progenitors and hexaploids. The phylogenetic tree differentiated *α*-gliadin sequences into three distinct groups representing *A* (black), *B* (green) and *D* (red) genomes.

analyzed to identify genome specific amino acid motifs (Table 3). Two short sequences (motifs) in NR1 and NR2 regions of *α*-gliadins differentiated most the amino acid sequences representing *A*, *B* and *D* genomes. One of the motifs (22 amino acid sequence) in NR1 domain positioned from 240 to 261 (Figure S1) distinguished all the 30 sequences representing *A* genome of diploids and 9 of the 11 sequences representing *A* genome in hexaploids. The motif also clearly differentiated 13 of 16 sequences of diploid (D genome) progenitors and 70 of 75 sequences representing *D* genome in hexaploids. All the *α*-gliadin sequences representing *A* genome of diploid (30) and hexaploid (11) could also be distinguished by 13 amino acid motif (PLGQGSFRPSQQN) in NR2 region. The 13 amino acid motif also distinguished 62 of 75 hexaploid sequences of *D* genome and 13 of 16 sequences of diploids.

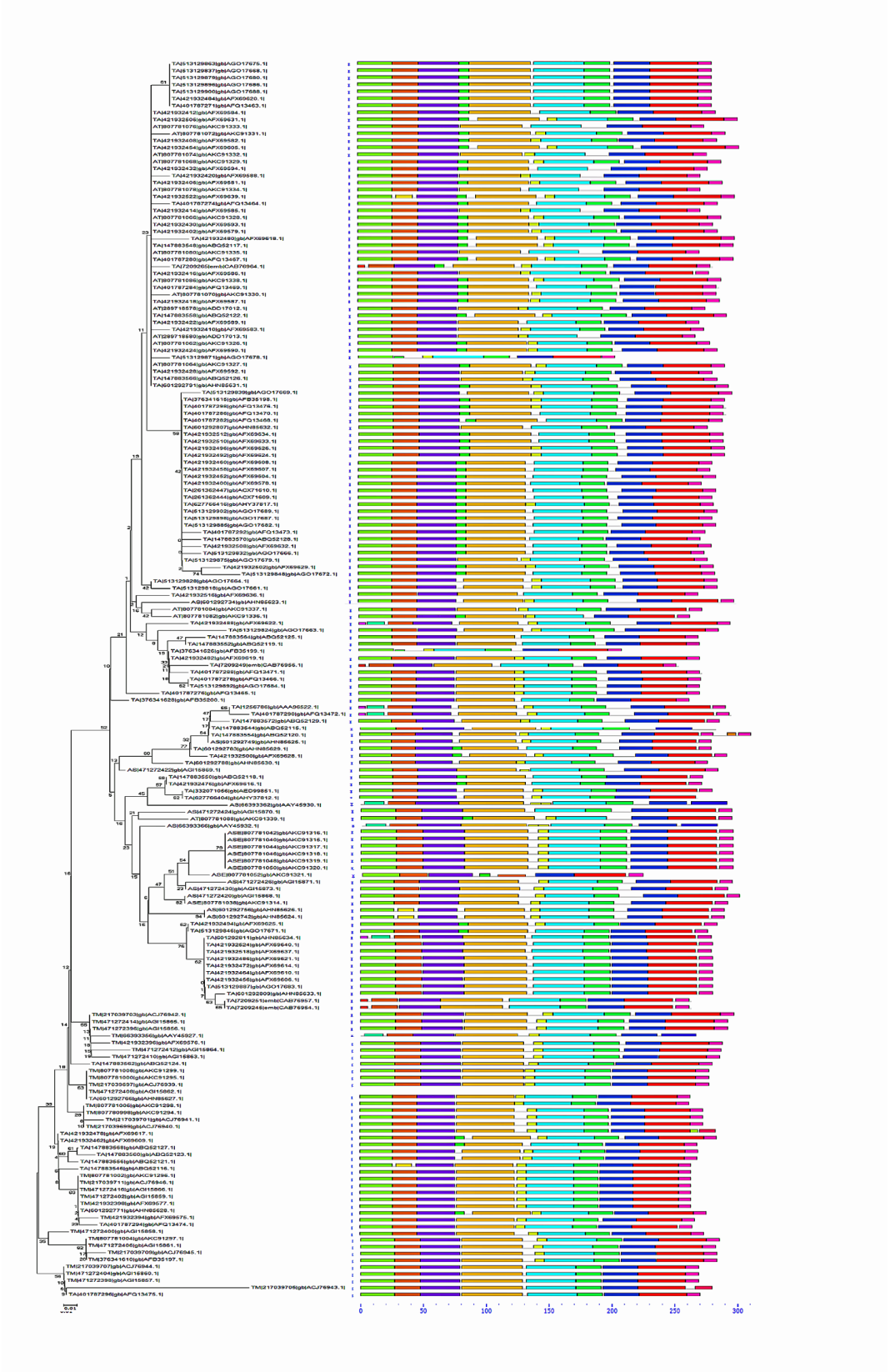

Figure 2. Schematic diagram of motif distribution of α-gliadins. Numbers written below each node are bootstrap values derived from 1,000 replicates. Twenty conserved novel motifs are shown with different colored boxes.

Sequence Polymorphism and Nucleotide Diversity of *α-Gliadin* Genes

Nucleotide diversity analysis of *α-gliadin* genes was conducted by Tajima *D* test (Tajima, 1989) using MEGA 6.0 software (Table 2). Large variations were observed among sequences of *α-gliadin* genes representing all the three genomes. Among diploid ancestors, *B* genome progenitors exhibited highest numbers of nucleotide polymorphic sites (ps= 0.31) followed by *A* genome (ps= 0.16) and *D* genome (ps= 0.13) progenitors. However, among hexaploids, *B* and *D* genome sequences exhibited higher number of nucleotide polymorphic sites (ps= 0.34 and 0.31) followed *A* genomes (ps= 0.19)**.** Highest Nucleotide diversity (P= 0.08) was observed in *B* genome donors (*Aegilopsspeltoides* and *Aegilopssearsii)* and *B* genome of hexaploids and lowest in *D* genomes of both diploids (0.02) and hexaploids (0.04). The higher nucleotide diversity in *B* genome was correlated with the reduced number of intact CD epitopes in *α*-gliadins and lower nucleotide diversity in *D* genomes with higher number of intact CD epitopes. *D* statistics further revealed significant deviations from neutrality distribution of *A* and *D* genome sequences of both diploids and hexaploids (P< 0.05). *D* genome showed highest *D* test value (-1.51) followed by *A* (-0.88) and B (-0.58) genomes of diploids. *D* test value in hexaploid sequences were -1.48 in *D*, -1.29 in *A* and 0.07 in *B* genomes (Table 2).

Table 2. Tajima *D* test statistics to identify nucleotide diversity of *α-gliadin* genes in *A, B* and *D* genomes of hexaploids wheat and their diploid progenitors.[a]

Species	Genome	m	n	S	ps	T	p	D
Diploid progenitors								
T. monococcum	A	30	248	40	0.16	0.041	0.03	-0.88
A. speltoides + A. searsii	B	19	167	52	0.31	0.089	0.08	-0.58
A. touschii	D	16	265	34	0.13	0.038	0.02	-1.51
Hexaploid wheat								
T. aestivum	A	11	274	52	0.19	0.065	0.05	-1.29
T. aestivum	B	26	242	75	0.31	0.081	0.08	0.07
T. aestivum	D	75	167	56	0.34	0.069	0.04	-1.48

[a] m= Number of sequences, n= Total Number of sites, S= Number of Segregating sites, ps= S/n, T= ps/a1, p= Nucleotide diversity and D is the Tajima test statistic.

Table 3. Amino acid motifs in NR1 and NR2 regions of *α*-gliadins differentiating them into three distinct groups representing *A, B* and *D* genomes.

Motif	Amino acid motif sequence	Number of *α*-gliadin sequences differentiated by 22 and 13 amino acid motif
22 Amino acid motif		
A Genome	*STYQLLQELCCQHLWQIPEQSQ*	30/30 (diploid) and 9/11 (hexaploids)
B Genome	*SSYQLLQQLCCQQLLQIPEQSR*	10/19 (diploid) and 12/26 (hexaploids)
D Genome	*STYQLVQQLCCQQLWQIPEQSR*	13/16 (diploid) and 70/75(hexaploids)
13 Amino acid motif		
A Genome	*PLGQGSFRPSQQN*	30/30 (diploid) and 11/11(hexaploids)
B Genome	*PSGQGSFQPSQQN*	16/19 (diploid) and 13/26 (hexaploids)
D Genome	*PSGQGSFQPSQQN*	13/16 (diploid) and 62/75 (hexaploids)

DISCUSSION

Identification and Frequency Distribution of CD Toxic Epitopes in *α*-Gliadins

There were large variations in the number of intact (T cell stimulatory) epitopes of *α*-gliadins representing *A, B,* and *D* genomes of bread wheat and its progenitors (Table 1). Intact Glia-α epitope was absent in *α*-gliadins encoded by *A* genome of both the diploid progenitors and hexaploid sequences. Intact Glia-α2 was absent in all the *α*-gliadins encoded by *A* and *B* genomes of diploids. However, both the Glia-αandGlia-α2 sequences were present in higher frequencies in *D* genome progenitors and encoded by *D* genomes of hexaploids. Intact Glia-α9 and Glia-α20 were present in high proportions in *A* and *D* while absent or rarely present in *B* genome encoded *α*-gliadins. Although intact Glia-α2 andGlia-α9 epitopes were absent in *B* genome diploid progenitors, 7 accessions in hexaploids exhibited both the epitopes in intact form in *α*-gliadins encoded by *B* genomes. Earlier reports indicated the absence of both Glia-α2 and Glia-α9 epitopes in all the diploid and hexaploids sequences (Van Herpen *et al.*, 2006). Overall, *α*-gliadinin *D* genome progenitors (*Aegilopstauschii)* and encoded by *D* genome of bread wheat exhibited all four intact CD epitopes. Several other reports also showed the presence of all epitopes in intact form in *α*-gliadins encoded by *D* genome (Van Herpen *et al.*, 2006; Li *et al.*, 2010; Xie *et al.*, 2010; Salentijn *et al.*, 2013; Li *et al.*, 2014a; Salentijn *et al.*, 2009; Van den Broeck *et al.*, 2010). Surprisingly, 5 accessions of hexaploids exhibited variant form of all the epitopes (none having intact epitopes) in *α*-gliadin encoded by *D* genomes. The accessions are TA|513129818|gb|AGO17661.1|; TA|147883564|gb|ABQ52125.1|; TA|421932488|gb|AFX69622.1|; TA|376341626|gb|AFB35199.1; and |TA|513129824|gb|AGO17663.1|. Since *D* genome in hexaploids is the major source of intact CD toxic epitopes, these accessions can be very useful source of improving wheat for reducing toxicity load among CD patients.

Phylogenetic Analysis and Assignment of *α*-Gliadin Sequences onto *A, B* and *D* Genomes

Both the diploid and hexaploid sequences of *α*-gliadins were phylogenetically clustered into 3 distinct groups representing *A, B,* and *D* genomes. Similarly, motif analysis showed three distinct groups of sequences represented by *A, B,* and *D* genomes. The characteristic feature of *α*-gliadins is the presence of two polyglutamine repetitive domains (QI & QII) and the size of which influences the visco-elastic properties of doughs. This is because of intermolecular interactions of the large numbers of glutamine side chains, which act as good hydrogen bond donors and acceptors (Masci *et al.*, 2000). Earlier reports indicated significant differences in the length of polyglutamine (QI and QII) repeats among all the three genomes, which could be used to distinguish genome specific sequences of *α*-gliadins (Van Herpen *et al.*, 2006; Xie *et al.*, 2010). However, in the present investigation, QI and QII repeat lengths could not distinguish *α*-gliadins representing different genomes. QII repeat length of *α*-gliadins showed significant differences between *α*-gliadins encoded by *A* and *B* between *B* and *D* genomes while no difference was observed between *A* and *D* genomes in diploids (Figure 3). However, in hexaploids, QII repeat length in *α*-gliadins was significantly different between *A* and *B* and between *A* and *D* genomes, however, no differences were observed between *B* and *D* genomes. As compared to diploids, QII repeat length of *B* genome was lower in hexaploids indicating loss of Q residues during early phase of evolution of hexaploid wheat (Figure 3).

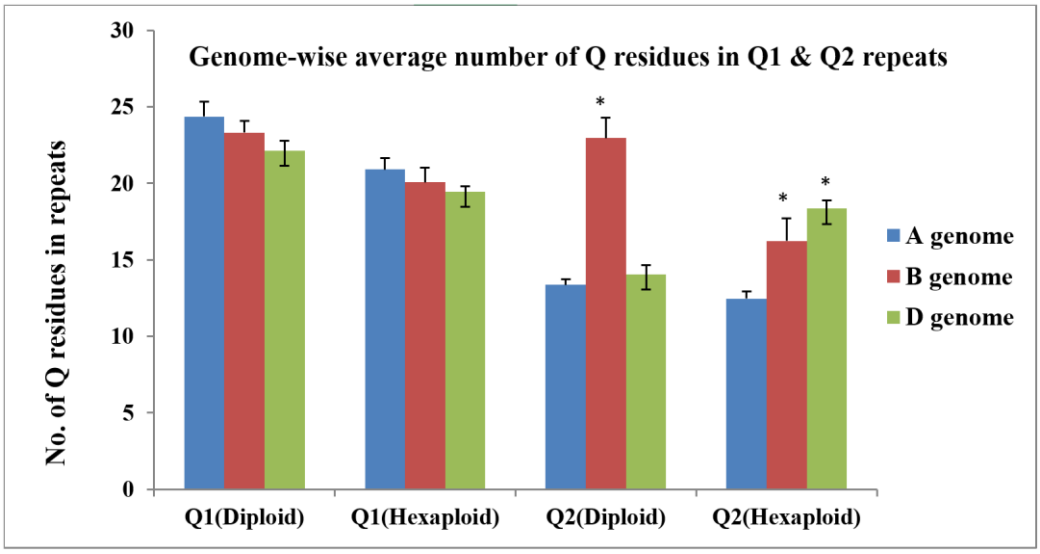

Figure 3. Genome wise average numbers of the glutamine residues in the first (QI) and second polyglutamine repeats (QII) of *α*-gliadins.

Since QI and QII length could not differentiate *α*-gliadin sequences encoded by different genomes, all the sequences were further analyzed to identify amino acid motifs distinguishing *α*-gliadin sequences representing *A*, *B*, and *D* genomes. Specific 22 and 13 amino acid motifs were found present in NR1 and NR2 domains, respectively. These motifs could differentiate *α*-gliadin sequences representing *A* and *D* genomes. The 22 amino acid motif distinguished all the 30 sequences representing *A* genome and 13 out of 16 sequences of *D* genomes of diploid progenitors. The motif also distinguished 70 of the 75 sequences representing *D* genomes and 9 out of 11 sequences representing *A* genome of hexaploids. Surprisingly, all the 70 sequences of *D* genome had valine at 245 positions and all the 9 sequences of *A* genome had glutamic acid (E), Histidine (H) and glutamine (Q) at 247, 252 and 261 positions, respectively. Similarly, 13 amino acid motifin NR2 region at 332-344 positions also differentiated all the *α*-gliadin sequences representing *A* genome of both diploids and hexaploids. Overall, both 22 and 13 amino acid motifs could distinguish *α*-gliadin sequences representing *A* and *D* genomes and the remaining sequences could be grouped into *B* genome. In hexaploid wheat, the donor of the *B* genome has been the most controversial and is still relatively unknown, in spite of a large number of attempts to identify the parental species (Huang *et al.*, 2002). This may be associated with the higher diversification rate of the *B* genome compared to the *A(u)* genome in the polyploid wheat (Petersen *et al.*, 2006), the incomplete chromosome pairing between B genome chromosomes and any diploid species (Blake *et al.*, 1999) and the fact that the B genome is relatively diverged from its putative diploid progenitors (Talbert *et al.*, 1995).

Nucleotide Diversity of *α-Gliadin* Genes

Tajima *D* test statistics was conducted to identify nucleotide diversity of *α-gliadin* genes. *B* genome donors (*Aegilopsspeltoides* and *Aegilopssearsii)* and *B* genome of hexaploids exhibited higher nucleotide diversity and reduced number of intact CD epitopes in *α*-gliadins as compared to *A* and *D* genome sequences. The statistics revealed significant deviations from neutrality distribution of *A* and *D* genome sequences of both diploids and hexaploids ($P< 0.05$). Highest D test value (negative) was exhibited by *D* genome sequences. The expectation for a neutral locus in a population is zero value of Tajima's *D*.

Positive values of Tajima's indicate paucity of rare alleles and preponderance of intermediate frequency alleles, while negative values exhibit a preponderance of rare alleles and a paucity of intermediate frequency alleles (Akhunov *et al.,* 2003). The negative values observed in Tajima's *D* test suggest that the *α-gliadin* gene has been subjected to purifying selection in *A* and *D* genomes of hexaploid and has not stabilized as the most potential allele of *α-gliadin* gene so far in genome evolution. The near to zero value (0.07) of Tajima's *D* in case of *B* genome of hexaploids demonstrated that variations are maintained in *B* genome sequences as per the mutation drift equilibrium. The positive value of Tajima's *D* in the *B* genome of hexaploid wheat is very likely due to strong subdivision of diploid *B* genome progenitor into two major sub-populations (*Ae. speltoides* & *Ae. searsii*). This sub-division has been acknowledged taxonomically by elevating individuals of the two sub-populations to sub-species. It is possible that *Ae. speltoides* is a significantly diverged form of the ancestral *B* genome donor (Blake *et al.*, 1999).

CONCLUSIONS

The information provided in the present study will be helpful in developing wheat safer for the celiac disease patient. Overall, the study exhibited large variations in *α-gliadin* gene sequences in diploid and hexaploid wheats. The set of 22 and 13 motifs in NR1 and NR2 regions of *α*-gliadin sequences, respectively, distinguished sequences representing *A* and *D* genomes, while *B* genome sequences exhibited wide variations. This will help in assigning new *α*-gliadin sequences onto *A, B* and *D* genomes. Some accessions were identified having *α*-gliadin with few or nil CD toxic epitopes useful in wheat improvement programs for management of celiac disease in human beings.

ACKNOWLEDGEMENTS

The financial assistance received from DBT, Government of India with the grant no. BT/PR5675/AGIII/103/863/2012 is gratefully acknowledged. The authors are thankful to the Director, Indian Institute of Wheat & Barley Research Institute (IIWBR), Karnal, for providing basic facilities. We also thank NCBI and various authors for making available sequence data of *α-gliadin* gene in various genomic backgrounds of bread wheat in the public domain.

REFERENCES

1. Akhunov, E. D., Akhunov, A. R., Linkiewicz, A. M., Dubcovsky, J., Hummel, D. and Lazo, G. 2003. Synteny Perturbations between Wheat Homoeologous Chromosomes by Locus Duplications and Deletions Correlate with Recombination Rates along Chromosome Arms. *Proc. Natl. Acad. Sci. USA,* **100**: 10836-10841.
2. Anderson, O. D., Dong, L., Huo, N. and Gu, Y. Q. 2012 A New Class of Wheat *Gliadin* Genes and Proteins. *PloS ONE,* **7(12):** 1-9.
3. Arentz-Hansen, H., Korner, R., Molberg, O., Quarsten, H., Vader, W. and Kooy, Y. M. C. 2000 The Intestinal T Cell Response to Alpha-Gliadin in Adult Celiac Disease is Focused on A Single Deamidated Glutamine Targeted By Tissue Trans Glutaminase. *J. Exp. Med.,* **191(4):** 603–612.
4. Blake, N. K., Lehfeldt, B. R., Lavin, M. and Talbert, L. E. 1999. Phylogenetic Reconstruction Based on Low Copy DNA Sequence Data in an Allopolyploid: The *B* Genome of Wheat. *Genome,* **42**: 351-360.
5. Chen, F., Xu, C., Chen, M., Wang, Y. and Xia, G. 2008. A New *A-Gliadin* Gene Family for Wheat Breeding, Somatic Introgression Line II-12 Derived from *Triticum aestivum and Agropyron elongatum*. *Mol. Breed.*, **22(4):** 675-685.
6. Gu, Y. Q., Crossman, C., Kong, X. Y., Luo, M. C., You, F. M. and Coleman-Derr, D. 2004. Genomic Organization of the

Complex *A-Gliadin* Gene Loci in Wheat. *Theor. Appl. Genet.*, **109(3):** 648-57.
7. Huang, S., Sirikhachornkit, A., Su, X., Faris, J. D., Gill, B. S. and Haselkorn, R. 2002. Genes Encoding Plastid Acetyl-Coa Carboxylase and 3-Phosphoglycerate Kinase of the *Triticum/Aegilops* Complex and the Evolutionary History of Polyploid Wheat. *Proc. Natl. Acad. Sci. USA*, **99**: 8133-8138.
8. Kawaura, K., Wu, J., Matsumoto, T., Kanamori, H., Katagiri, S. and Ogihara, Y. 2012. Genome Change in Wheat Observed through the Structure and Expression of *A/B-Gliadin* Genes. *Funct. Integr. Genomic*, **12**: 341–355.
9. Li, G. R., Lang, T., Yang, E. N., Liu, C. and Yang, Z. J. 2014a. Characterization and Phylogenetic Analysis of *A-Gliadin* Gene Sequences Reveals Significant Genomic Divergence in *Triticeae* Species. *J. Genet.*, **93**: 725–31
10. Li, G., Zhang, T., Ban, Y. and Yang, Z. 2010. Molecular Characterization and Evolutionary Analysis of A-Gliadin Genes from *Eremopyrumbon aepartis* (*Triticeae*). *J. Agric. Sci.*, **2(4):** 30-36.
11. Li Y., Xin R., Zhang D. and Li S. 2014b. Molecular Characterization of *A-Gliadin* Genes from Common Wheat Cultivar Zhengmai 004 and Their Role in Quality and Celiac Disease. *Crop J.*, **2**: 10–21.
12. Masci S., D'Ovidio R., Lafiandra D. and Kasarda D. D. 2000. A 1B Coded Low-Molecular-Weight Glutenin Subunit Associated with Quality in Durum Wheats Show Strong Similarity to Subunits Present in Some Bread Wheat Cultivars. *Theor. Appl. Genet.*, **100**: 396–400.
13. Mc Manus, R. and Kelleher, D. 2003. Celiac Disease-The Villain Unmasked? *New Engl. J. Med.*, **348**: 2573–2574.
14. Molberg, O., Solheim, F. N., Jensen, T., Lundin, K. E., Arentz-Hansen, H., Anderson, O. D. Uhlen, A., and Sollid L. M., 2003. Intestinal T-Cell Responses to High Molecular Weight Glutenins in Celiac Disease. *Gastroenterology*, **125**: 337–344.
15. Molberg, Ø., Uhlen, A. K., Jensen, T., Flæte, N. S., Fleckenstein, B., Arentz-Hansen, H., Raki, M., Lundin, K. E. A., and Sollid, L. M., 2005. Mapping of Gluten T Cell Epitopes in The Bread Wheat Ancestors: Implications for Celiac Disease. *Gastroenterology*, **128**: 393–401.
16. Paulsen, G., Lundin, K. E., Gjertsen, H. A., Hansen, T., Sollid L. M. and Thorsby, E. 1995. HLA-DQ2-Restricted T-Cell Recognition of Gluten-Derived Peptides in Celiac Disease. Influence of Amino Acid Substitutions in the Membrane Distal Domain of DQ Beta 1*0201. *Hum. Immunol.*, **42**: 145–153.
17. Petersen, G., Seberg, O., Yde, M. and Berthelsen, K. 2006. Phylogenetic Relationships of *Triticum* and *Aegilops* and Evidence for the Origin of the *A*, *B*, and *D* Genomes of Common Wheat (*Triticumaestivum*). *Mol. Phylogenet. Evol.*, **39**: 70-82.
18. Qi, P. F., Chen, Q., Ouellet, T., Wang, Z., Le, C. X., Wei, M., Wei, Y. M., Lan, X. Jin., and Zheng, Y. L., 2013 The Molecular Diversity of *A-Gliadin* Genes in The Tribe *Triticeae*. *Genetica*, **141**: 303–310.
19. Salentijn, E. M. J., Goryunova, S. V., Bas, N., Van den Meer, I. M., Van den Broeck, H., Bastien, T. Gilissen, L. J. W. J., and Smulders, M. J. M., 2009. Tetraploid and Hexaploid Wheat Varieties Reveal Large Differences in Expression of Alpha-Gliadins from Homoeologous Gli-2 Loci. *BMC Genom.*, **10** : 10-48
20. Salentijn, E. M. J., Esselink, D. G., Goryunova, S. V., Vanden Meer, I. M., Gilissen, L. J. W. J. and Smulders, M. J. M. 2013. Quantitative and Qualitative Differences in Celiac Disease Epitopes among Durum Wheat Varieties Identified Through Deep RNA-Amplicon Sequencing. *BMC Genomics*, **14**: 1-16.
21. Shewry, P. R. and Halford, N. G. 2002. Cereal Seed Storage Proteins Structures, Properties and Role in Grain Utilization. *J. Exp. Bot.*, **53**: 947–958.
22. Shewry, P. R. and Tatham, A. S. 2016. Improving Wheat to Remove Coeliac Epitopes but Retain Functionality. *J. Cereal Sci.*, **67**: 12-21.
23. Sjostrom, H., Lundin, K. E., Molberg O., Korner R., McAdam, S. N., Anthonsen, D. Quarsten, H., Nore, O., Roepstorff, N. P., Thorsby, E., and Sollid L. M., 1998. Identification of A-Gliadin T-Cell Epitope in Coeliac Disease: General Importance of Gliadin Deamidation for Intestinal T-Cell Recognition. *Scand. J. Immunol.*, **48:** 111–115.
24. Spaenij-Dekking, L., Kooy-Winkelaar, Y., Van Veelen, P., Drijfhout, J. W., Jonker,

H., Van Soest, L. *et al.* 2005. Natural Variation in Toxicity of Wheat: Potential for Selection of Nontoxic Varieties for Celiac Disease Patients. *Gastroenterology*, **129**: 797–806.

25. Tajima F. 1989. Statistical Method for Testing the Neutral Mutation Hypothesis b by DNA Polymorphism. *Genetics*, **123:** 585–595.
26. Talbert, L. E., Blake, N. K., Storlie, E. W. and Lavin, M. 1995. Variability in Wheat Based on Low-Copy DNA Sequence Comparisons. *Genome*, **38**:951-57.
27. Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. 2013. MEGA 6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.*, **30**:2725-2729.
28. Thompson, J. D., Higgins, D. G. and Gibson, T. J. 1994. CLUSTAL W: Improving the Sensitivity of Progressive Sequence Alignment through Sequence Weighting, Position-Specific Gaps Penalties and Weight Matrix Choice. *Nucl. Acids Res.*, **22(22):** 4673–4680.
29. Timothy, L., Bailey, L. and Charles, E. 1994. Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, San Diego, California, USA. **2**: 28-36.
30. Vaccino, P., Becker, H. A., Brandolini, A., Salamini, F. and Kilian, B. 2009. A Catalogue of *Triticummonococcum* Genes Encoding Toxic and Immunogenic Peptides for Celiac Disease Patients. *Mol. Genet. Genomics*, **281**: 289–300.
31. Vader, W., Stepniak, D., Kooy, Y., Mearin, L., Thompson, A., van Rood, J. J. Spaenij, L., and Koning, F., 2003. The HLA-DQ2 Gene Dose Effect in Celiac Disease is Directly Related to the Magnitude and Breadth of Gluten-Specific T Cell Responses. *Proc. Natl. Acad. Sci.* USA, **100**: 12390–12395.
32. Van den Broeck, H., Hongbing, C., Lacaze, X., Dusautoir, J. C., Gilissen, L., Smulders, M. and Meer, I. V., 2010. In Search of Tetraploid Wheat Accessions Reduced in Celiac Disease-Related Gluten Epitopes. *Mol. Biosyst.*, **6**: 213-220.
33. Van Herpen, T. W. J. M., Goryunova, S. V., van der Schoot, J., Mitreva, M., Salentijn, E., Vorst, O. Schenk, M.F., Veelen, P. A.V., Koning, F., Soest, L. J. M. V., Vosman, B., Bosch, D., Hamer, R. J., Gilissen, L. J. W. J., and Smulders, M. J. M., 2006. *Alpha-Gliadin* Genes From the *A, B,* and *D* Genomes of Wheat Contain Different Sets Of Celiac Disease Epitopes. *BMC Genomics*, **7:** 1. doi:10.1186/1471-2164-7-1.
34. Vande Wal, Y., Kooy, Y. M., van Veelen, P. A., Pena, S. A., Mearin, L. M., Molberg O. *et al.* 1998. Small Intestinal T Cells of Celiac Disease Patients Recognize A Natural Pepsin Fragment of Gliadin. *Proc. Natl. Acad. Sci. USA*, 95: 10050–10054.
35. Xie, Z., Wang, C., Wang, K., Wang, S., Li, X., Zhang, Z., Ma, W., and Yan, Y., 2010. Molecular Characterization of the Celiac Disease Epitope Domains in A-Gliadin Genes In *Aegilopstauschii* and Hexaploids Wheats (*Triticumaestivum* L.). *Theor. Appl. Genet.*, **121**: 1239–1251.

شناسایی توالی موتیف های ویژه ژنوم در α-Gliadins و نمونه های ثبت شده گندم با اپی توپ های بیماری سلیاک

س. سینگ، س. رام، و س. ناروال

چکیده

در میان Gliadin ها، α-Gliadins به خاطر حضور اپی توپ های سمی، پروتئین های فعال ومهمی در شروع بیماری سلیاک در افراد بشر هستند. در این پژوهش، مجموعه ای شامل ۱۷۷ توالی ژنی α-Gliadin و پروتئین های متناظر آن ها تجزیه شد. در نتیجه، ۲۰ نمونه ثبت شده هگزاپلویدی شامل ۱، ۱۴، و ۵، به ترتیب نماینده ژنوم A،B ، و D بدون CD-epitopes های سالم و دست نخورده در α-Gliadin شناسایی شد. سپس، ۲۲ و ۱۳ موتیف حفاظت شده (conserved motifs) به ترتیب در دامنه های غیر تکراری NR1و NR2 از α-Gliadin، همه توالی های آمینو اسیدهای رمز گذاری شده با ژنوم A در دیپلوید ها و هگزاپلوید ها را تمایز یابی کردند. بیشتر توالی آمینو اسیدهای رمز گذاری شده با ژنوم D (۷۰ تا از ۷۵ هگزا پلوید و ۱۳ تا از ۱۶ دیپلوید) با ۲۲ موتیف آمینواسید قابل شناسایی بود. مشاهدات حاکی از تغییرات زیاد و تعداد کمتر CD-epitopes های سالم و دست نخورده در α-Gliadin مربوط به ژنوم B بود. در مقایسه با دیپلوید ها، طول تکرار(repeat length) دامنه تکرار شوندهQII پلیگلوتامین ژنومB درهگزا پلوید ها کمتر بود و این امر به ازدست رفتن بقایایQ در طی تکامل گندم هگزا پلوید اشارت داشت. از این اطلاعات می توان در تخصیص هر توالی α-Gliadin روی ژنوم های A، B، و Dو شناسایی نمونه های ثبت شده گندم با CD-epitopesکمتر بهره جست. نتایجی که در اینجا ارایه شده می تواند برای برنامه های اصلاح و بهبود گندم با هدف مدیریت کردن بیماری سلیاک در افراد بشر مفید باشد.