Australian Journal of Crop Science AJCS
AJCS 9(12):1131-1138 (2015) ISSN:1835-2707

Single nucleotide polymorphism in wheat triticin gene (*Tri 1*) among diverse wild wheat species

Sonia Goel[1], Manoj Nath, Nisha Singh, Sapna Grewal, Shilpa D'souza, Ajay Mahato, N. K. Singh*

National Research Centre on Plant Biotechnology, Pusa campus, New Delhi 110012, India

***Corresponding author: nksingh@nrcpb.org**

Abstract

Wheat is one of the major sources of human food and nutrition it is rich in protein but limiting in essential amino acids like lysine, tryptophan and threonine. Hence, there is a need for nutritional quality improvement of wheat using modern functional genomics tools. Here, we analyzed single nucleotide polymorphism (SNPs) in high nutritional quality triticin gene (EU482413) among 21 wild wheat accessions using a targeted re-sequencing approach. Overall, 91 SNPs were found in the intronic and the hypervariable region (HVR) of the triticin gene of these 86 SNP/indels were found in the three introns and only 5 SNPs in the HVR region, probably due to strong selection pressure on the coding sequences. Wheat accession, *T. monococcum* 2802 showed one extra lysine residue in the HVR region. The reference sequence of the triticin was found to be similar to D genome accessions. These SNP data sets were used in phylogenetic and PCA analysis to assess the sequence diversity in these varieties. The wild wheat accessions were separated into three clusters corresponding to their genome. This study can be of immense use in finding novel variants of triticin to enhance the nutritional quality of wheat in future

Keywords: Agarose gel electrophoresis; Hypervariable region; SNP detection; Triticin; Wheat.
Abbreviations: HVR_Hypervariable region, IARI_ Indian Agriculture Research Institute, PAU_Punjab Agricultural University, SNP_Single Nucleotide Polymorphism.

Introduction

The primary source of all protein in the human diet are majorly contributed by legumes and cereals. Cereals contribute about 50% of the per capita energy intake worldwide and 65% in developing countries. Though the protein content of cereals is lesser (7-14%) than that of legumes (20-40%), they account for about 45% of the daily per capita protein supply in the world and approximately 63% in developing countries. This data clearly illustrates the importance of cereals in our daily life, but the poor nutritional quality and deficiency of cereals in some essential amino-acids still remains one of the major concerns. Lysine is the first nutritionally limiting essential amino acid in most cereals; second being tryptophan in maize and threonine in other cereals (Eggum and Beames, 1983). Plant geneticists and breeders are making continuous efforts to improve the quality of cereal seed proteins. Recent genetic, molecular and biochemical evidence suggests that lysine synthesis and catabolism are regulated by complex mechanisms. The idea of producing high lysine mutants was raised years ago but apart from the high quality protein maize lines currently commercially available, nothing much has been achieved (Ferreira 2005). The production of mutants is not an easy task in wheat due to its polyploid nature, but efforts are being made to enhance its nutritional status *via* various other methods. Wheat is one of the most important staple diet consumed by a majority of the global population. It is the single largest source of protein and second only to rice in fulfilling the daily calorie need of Indian population. Target for enhancing grain quality of wheat includes increasing the proportion of essential amino acids like lysine, improving bread-making quality by manipulating high molecular glutenin proteins and modifying starch composition (Patnaik and Khurana, 2001). Successful efforts have been made to improve the bread making quality by exploiting high molecular weight glutenin subunits (HMW-GS) of wheat seed proteins (Alvarez et al., 2000) but increasing the lysine content, which is one of the most limiting amino acids in wheat, is still a challenge. This challenge can be overcome by enhancing triticin protein level in wheat. Triticin is a minor storage protein of wheat endosperm with good essential amino- acid balance, but it contributes less than 5% of the total seed proteins. It has lysine rich decapeptide repeat motif in its hypervariable region (HVR). This protein has been characterized well at genetic and biochemical levels (Singh et al., 1988). The genetic manipulation of triticin gene in increasing lysine content can be made possible either by enhancing the expression using more efficient seed storage protein promoters or manipulating the hypervariable region. Presence of a lysine-rich repetitive domain in the hyper variable region of triticin gene offers new opportunities for the genetic engineering of triticin for even higher lysine content (Singh et al., 1993). Hypervariable region lies in the third exon part of triticin gene comprising 100 amino acids and contains 10 units of lysine (Singh et al., 2009). Wild wheat progenitors have always proved to be a source of the wide variation in useful traits and are well explored for traits like drought tolerance where they have been proved to be the best source (Budak, 2013). Seed storage proteins were also investigated in different wild wheat varieties and found that the wild progenitors are important genetic resources. Wild

emmer wheat (*Triticum turgidum* ssp. *dicoccoides*), the progenitor of domesticated wheat, was tested for genetic diversity studies regarding grain nutrient concentrations by Merav Chatzav in 2010 and he found that the concentrations of grain zinc, iron and protein in wild accession was about two-fold higher than in the domesticated genotypes. Some work based on diversity analysis using SNP genotyping has also been done much recently on wild emmer wheats (Jing Ren, 2013). These examples are wide enough to support our view on the importance of wild and exotic species of wheat.

Single nucleotide polymorphisms (SNPs) are the predominant class of genetic variation present in plant species and can be used as molecular markers. They are usually bi-allelic and occur at high density in the genome. Occurrence of SNPs in non-coding regions is more frequent as compared to the coding region of the genome, due to some natural selection process acting on the genome. Beside this, some other factors like recombination and rate of mutation can also detriment the SNP density in the genome, (Ramakrishna et al., 2002). SNPs occuring in the coding region enables the assessment of polymorphisms potentially directly affecting the phenotype (Syvanen, 2001). In contrast intronic SNPs can likely influence splicing of the mRNA of the given gene. The mRNA may undergo alternative splicing and the intron may get included into alternative forms of protein. This SNP may thereby affect the structure and function of coded protein. Moreover, introns may have their own transcription units producing regulatory RNAs or perhaps small proteins in that cases SNP may affect the function of the corresponding products of intron expression. SNPs at intronic region may be useful in association studies in wheat and its wild relative germplasm. Since the introns positions in wheat genes are highly conserved between the genes (Dubcovsky et al., 2001; Ramakrishna et al., 2002), intron-based sequence differences would be quite useful for genetic analysis and wheat breeding program in hexaploid wheat (Hwayoung, 2013). In the present study, triticin protein has been explored in different wild wheat progenitors to study the variation on DNA as well as on sequence level. Detection of SNPs in the *VRN-A1* gene has already been exploited in wheat to find genotypic and phenotypic diversity (Liang 2011). Exotic wheat varieties with improved triticin protein can be used in any selection strategies for breeding programes. In the present study, triticin gene was analyzed in wheat and its wild progenitors through PCR amplicon based sequencing by Sanger's dideoxy method on automated sequencer (MegaBACE 4000). Further SNPs, insertion/deletion (indels) were discovered from *de novo* assembly data followed by cluster and phylogenetic analysis.

Results

PCR Amplification of intron and HVR region of triticin

Triticin gene was amplified in twenty-one wild wheat accessions with intron and HVR primers and analyzed by agarose gel electrophoresis. Ta_triticin_int1, Ta_triticin_int2 and Ta_triticin_int3 primer of intronic regions showed the expected size band of 210 bp, 190 bp and 110 bp respectively with no polymorphism on agarose gel (Fig 1A). NKS_triticin 1 primer was used to amplify HVR and the results were analyzed on 4% agarose gel where *A. squarrosa* showed a specific band of 300 bp size while *T. monococcum, T. umbellulata* and *A. speltoides* showed a slightly lower band of lesser base pairs.

Sequence variation in the Intron and HVR regions of triticin

The amplified PCR product of all the DNA samples were sequenced after purification and aligned with phred phrep to obtain the contigs. Sequences were aligned using ClustalW (Thopmson et al., 1994) and BioEdit (Hall, 1999) after trimming the primer part. A total of 56 SNPs and 7 indels were observed in Intron1 region (Table 3). Wild wheats having D genome were aligned as one group against reference sequence of Chinese Spring whereas; wheat samples with other diverse genomes (A, B, SYand U) formed another separate group (Fig 2A, B). *T. monococcum* (A genome) accessions showed 43 SNPs and *A. speltoides* (B genome) showed 40 SNPs in Intron1 against reference sequence. *T. monococcum* WLT2804 showed high similarity to *A. speltoides* in case of SNPs, which conclude the near relation of A and B genomes. *A. umbellulata* showed only one indel in Intron1 region and was found to be highly similar to Chinese Spring sequence. Elymus was the only accession of genome SY which showed long indels in intron 1. Least sequence variation was found among accessions of D genome wheat and hence, showed the maximum similarity with Chinese Spring reference sequence. D genome species *T. tauschii* showed six SNPs and *A. squarrosa* didn't show any SNP in intron1 region.

In Intron2 region, *A. speltoides* and Elymus showed very similar SNPs as four out of five Elymus SNPs were common to that of *A. speltoides*. *T. monococcum* 2802 showed two SNPs and *T. umbellulata* showed six SNPs in Intron2 region. Chinese Spring was again found to be most similar to *A. squarrosa* accession having D genome (Fig 3). A total of 15 SNPs and 1 indel was found in Intron3 (Table 3). All the wild wheat accessions showed A instead of G at 1629th position *T. monococcum* showed maximum of six SNPs in this region while D genome *A. squarrosa* showed nine SNPs in this region and again found to be similar to Chinese Spring reference sequence (Fig 4). HVR is present in exon 3 region of the triticin gene. We amplified seven wild wheat accessions for the HVR region including three accessions of *A. squarrosa* (D genome), one *T. monococcum* (A genome), one *T. umbellulata* (U genome), one *A. speltoides* (B genome) and one *T. tauschii* (D genome) (Table 4). A total of five SNPs were found, two common in all seven accessions and one specific to *T. monococcum* 2802, *A. squarrosa* 14091 and *A. squarrosa* 3758 each. The consensus sequences were converted into amino acid sequences and aligned with reference Chinese Spring triticin sequence which revealed the presence of histidine instead of glutamic acid in *A. squarrosa* 14091 and *A. squarrosa* 3758 at 291th position And Lysine instead of methionine in *T. monococcum* 2802 at 295th position of the HVR region. All the seven accessions revealed proline to serine transition at 298th position and glutamic acid to glutamine transition at 353rd position (Fig 5B). Position of SNP was assisgned from starting codon.

Diversity and phylogenetic analysis of the triticin sequence in wild and cultivated wheat

Principal component analysis (PCA) summarizes the major patterns of variation in an SNP data set from Intron1, Intron2 and HVR in eleven wheat varieties (presence of SNPs) were performed. The cluster analysis, PCoA based on the SNP data, revealed three distinct clusters for all the varieties related to their genome type. The results indicated that there were obvious genetic variations between cultivars based on their genome PCoA and UPGMA dendrogram, which were

Table 1. List of 21 wild wheat cultivars used in the study.

S.No.	Genome	Accession	S.No	Genome	Accession
1	A	*T. monococcum WLT2804*	12	D	*A. squarrosa 14114*
2	A	*T. monococcum 2802*	13	D	*A. squarrosa 9810*
3	A	*T. monococcumaegilopside WLT2800*	14	D	*A. squarrosa 14197*
4	B	*A. speltoides 348557*	15	D	*A. squarrosa 9788*
5	B	*A. speltoides var linguistica 1902*	16	D	*A. squarrosa 14230*
6	U	*A. umbellulata*	17	D	*A. squarrosa 43472*
7	SY	*Elymus*	18	D	*A. squarrosa 14163*
8	D	*A. squarrosa 9816*	19	D	*A. squarrosa 14231*
9	D	*A. squarrosa 3759*	20	D	*A. squarrosa 500516*
10	D	*A. squarrosa 3787*	21	D	*T. tauschii 14222*
11	D	*A. squarrosa 14091*			

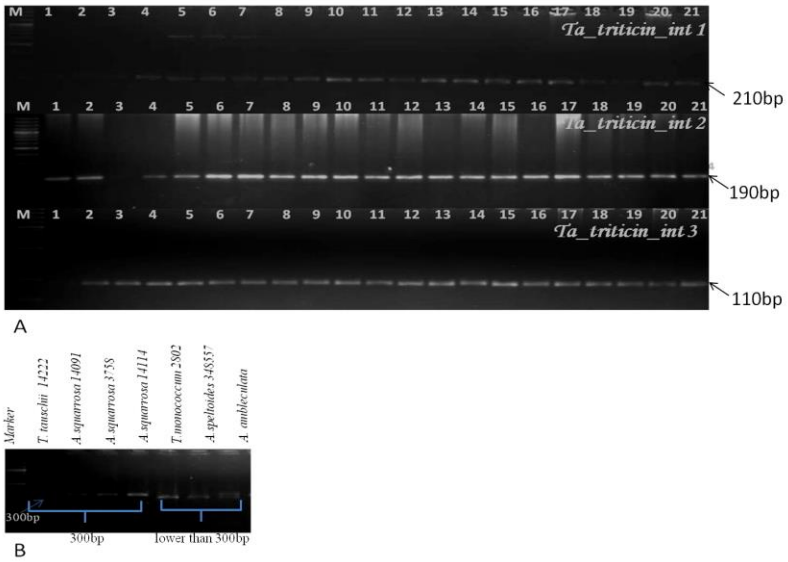

Fig 1. (A) PCR amplification of Triticin gene (Intron 1, 2 and 3 regions) among 21 wild wheat accessions using primer Ta_triticin_int1, 2, 3. (B) PCR amplification of Triticin gene HVR among seven wild wheat accessions.

based on retrieved SNP from eleven wheat varieties. The first and second principal coordinates explained 73.24% and 12.74% of the total variation, respectively (Fig 6). The overall three principal coordinates explained total 91.72 % variation in wheat varieties. The PCA analysis separated the entire D genome varieties (*A. squarrossa* 14091, *A. squarrossa* 3758, *A. squarrossa* 14114 and *T. tauschii* 1422) assembled in one separate cluster and it well correspond to UPGMA dendrogram (Fig 7). Three varieties of A genome (*T. monococcum* WLT2804, *T. monococcum* 2802 and *T. monococcum aegilopside* WLT2800) along with one variety of B genome (*A. speltoides linguistica*1902) were separated in second cluster. In third cluster consists of mixture of three types of genome B (*A. speltoides* 348557), SY (Elymus) and U (*A. umbellulata*) genome. The UPGMA clustering based on CSchord 1967 method (Cavalli-Sforza LL et al., 1967) genetic distance grouped wheat varieties into three distinct groups (Fig 7). The UPGMA tree revealed that wheat varieties clustered into smaller sub-groups based on the type of genome. For example D genome varieties clustered together (*A. squarrossa* 14091, *A. squarrossa* 3758, *A. squarrossa* 14114 and *T. tauschii* 1422). In second group was of U (*A. umbellulata*) genome in company with two A genome (*T. monococcum aegilopside WLT2800* and *T. monococcum 2802*). In third group Elymus along with two B (*A. speltoides 348557, A. speltoides var linguistica 1902*) genome varieties and one A genome (*T. monococcum* WLT2804) formed a separate group. The PCA analysis supports the grouping as observed by UPGMA based dendrogram. UPGMA and PCA based clustering and grouping of wheat varieties is consistent with the classification of genome types. SNPs and indels were observed in different wheat accessions with wide genome A, B, D, U and SY in intron and HVR regions. One major observation found here was Chinese Spring sequence similarity with D genome accession, *A. squarrosa* in all the three introns as well as HVR region. Chinese Spring is hexaploid in nature with ABD genome and its triticin sequence is available at NCBI, but till date, no such evidence was obtained regarding the available triticin sequence of Chinese Spring. Here the collaborative study of different genomes showed the resemblance of triticin gene most with *A. squarrosa* accessions which supports our conclusion that triticin sequence of Chinese Spring taken from NCBI is a result of Chinese Spring D genome sequence out of ABD.

Discussion

Wheat is the most consumed food crop, but it has poor balance of amino acids due to lack of lysine and threonine. Focus was raised on increasing the level of lysine in wheat to improve its nutritional value. Target protein was thought to be triticin which offers new opportunities to increase the level of lysine in wheat seeds to improve its nutritional quality. The HVR of triticin, which is over 100 amino acids long, can accommodate more than ten of these lysine rich decapeptide repeat units. Since this repeat motif is naturally present in the wheat triticin, it is unlikely to adversely affect

Table 2. Details of primers for intron and HVR region amplification designed from triticin sequence of Chinese Spring wheat variety.

S. no.	Primer Name	Sequence (5'-3')	Expected size of amplicon(bp)
		Intron primers	
1	Ta_triticin_int1 forward	GCGATACCACAACACTCATGG	210
	Ta_triticin_int1 reverse	GCATCCAGGAAAAGACAGTCC	
2	Ta_triticin_int2 forward	TTTTCGACGTAAACAACTATGC	190
	Ta_triticin_int2 reverse	CCAGCGAACAAAAATTCCTAC	
3	Ta_triticin_int3 forward	TCTCAACTCCCAAACGTTCC	110
	Ta_triticin_int3 reverse	GGCGTTCTGCGATCATAAAT	
		HVR primer	
4	NKS_triticin1forward	CCCTTAAATTTCTGAAGCCTGTT	300
	NKS_triticin1Reverse	CTGCCTGAGATTGTTCTTGTTG	

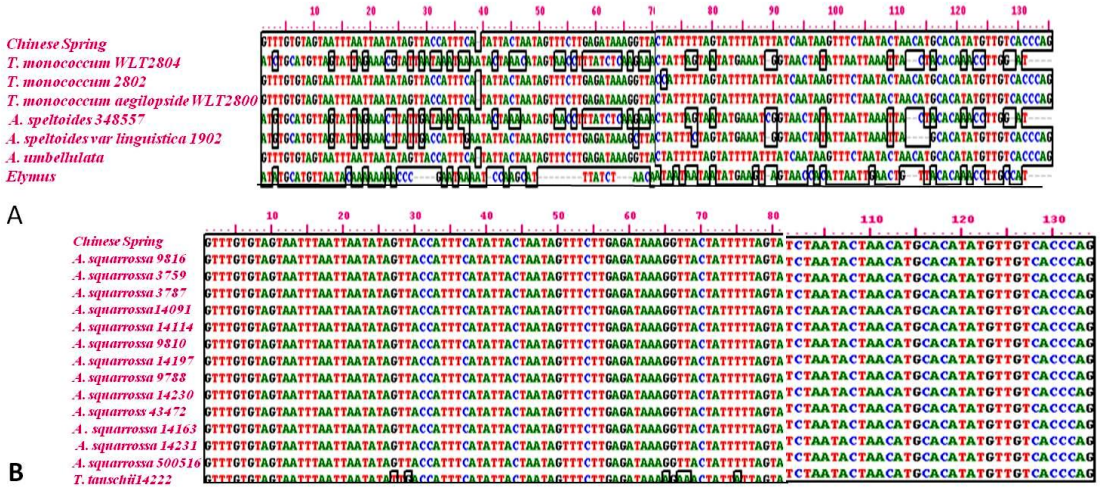

Fig 2. Alignment of the sequences of triticin intron1 region from wild wheat accessions along with triticin of Chinese Spring. A. Alignment of wild wheat accessions with diverse genomes A, B, SY, U.

the biological stability of triticin in the seeds, or the normal kernel development (Singh et al., 1993). Detailed work on HVR of triticin and variation on sequence level was necessary to get further knowledge of this important region from quality point of view. Here we have analyzed different wild wheat progenitors of diverse five wheat genomes A, B, D, U and SY for estimate variation in triticin gene. Intron1, 2 and 3 and HVR of triticin gene were amplified and aligned in different wheat accessions against Chinese Spring reference sequence to observe SNPs. Earlier, also (Shailaja et al., 2002). Analyzed wheat progenitors like *A. speltoides*, *A. longissima*, *A. squarrosa*, *T. urartu* along with bread varieties HD-2329, UP-262 and Kalyansona for natural variation in the hypervariable region. They used different sets of primers and optimized PCR conditions for amplification. They concluded the expression level of *the Tri-1D gene* to be higher than *Tri-1A* gene, but no focus was given on sequence variation. Monica et. al. also attempted to amplify triticin gene in diploid wheat, but no significant polymorphism was observed which can be thought to affect the lysine content (Monica et al., 2007). Till date, no such study has been conducted to find out SNP variation in triticin gene. Here we have designed primers for the amplification of introns as well as HVR and carried out sequencing of PCR product to detect SNPs and Indels in diverse wheat progenitors. In our study of finding out SNPs in triticin gene among wild wheat accessions, we observed no polymorphism in intron region on 4%agarose gel however three accessions *T. monococcum* (A genome), *T. umbellulata* (U genome), *A. speltoides* (B genome) showed a lower band of less than 300 bp size when amplified for HVR, this probably due to sequence level difference of A, B, D and U genome. Further polymorphism was studied at sequencing level. Sequencing of the accessions showed that Elymus has long indels in Intron1 region. Elymus is an ancient wheat variety tetraploid (SY) in nature thought to be important as it is resistant to a devastating disease scab (Wang et al., 1999). So it can be explored more for other factors like quality improvement. Elymus and *A. speltoides* have common SNPs in Intron1 and Intron2 regions (Fig 2A, 3A). Earlier also B genome and S genome were found to be close relative or S genome as a derivative of B genome (Haider 2013). Elymus and *A. Speltoides* accessions fall into the same branch in the UPGMA diagram (Fig 7). Highest SNP variation was observed in Intron1 region out of all the three introns with a total of 50 SNPs and 6 indels in intron1 while intron2 and intron3 showed 15 SNPs each. One accession of A genome *i.e. T. monococcum* 2802 was found to have similar SNPs as that of *A. speltoides* in intron 1. The positioning of SNPs was done from the starting codon of the triticin gene. We have used a range of five genomes of wheat to find variation among different species and also we need to focus on intron regions as well and where we observed remarkable SNPs. *A. umbellulata* having U genome showed no SNP in intron1 region but 4 SNPs in intron2 region, this observation placed U genome *A. umbellulata* near to D genome accessions as they also showed no SNPs in Intron1 region. Earlier, Gandhi et al. (2005) also concluded the near relation of U genome to D than to the A and B genomes (Gandhi et al., 2005).

Table 3. List of SNPs and indels in the intron and hypervariable region of wheat triticin gene among wild wheat accessions taking Chinese Spring as reference sequence.

	No. of bases sequenced	No. of SNPs	No. of In/dels	Synonymus or non Synonymus SNPs (s/ns)	Amino Acid alteration	
					Position	AAchange
*Intron 1	130	56	7	s	-	
*Intron 2	120	15	1	s	-	
*Intron3	90	15	1	s	-	
**HVR (Exon)	270	5		4 ns	24	H/E
					28	M/K
					31	P/S
					86	E/Q

* Non coding region of twenty one wild wheat accessions. ** Coding region from seven wild wheat accessions.

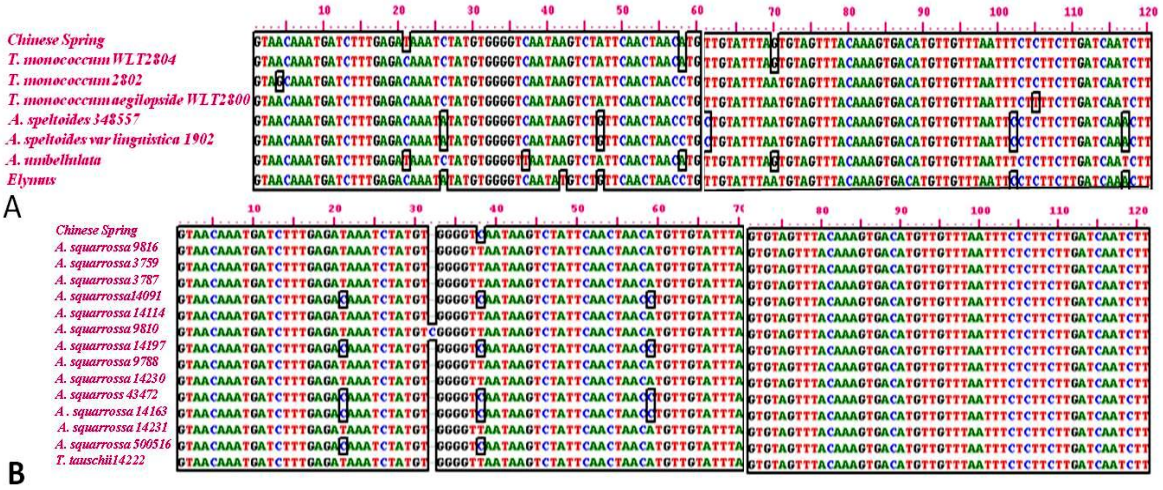

Fig 3. Alignment of the sequences of triticin intron2 region from wild wheat accessions along with triticin of Chinese Spring.

Table 4. List of wild wheat accessions used for amplification of hypervariable region of triticin.

S.no.	Accessions	Genome
1	*T. tauschii 14222*	D
2	*A. squarrosa 14091*	D
3	*A. squarrosa* 3758	D
4	*A. squarrosa 14114*	D
5	*T. monococcum 2802*	A
6	*A. speltoides 348557*	B
7	*A. umbellulata*	U

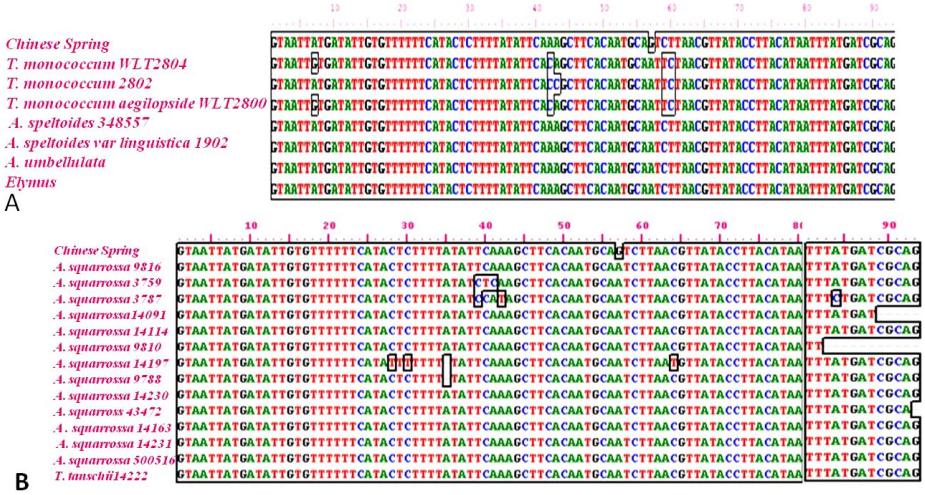

Fig 4. Alignment of the sequences of triticin intron3 region from wild wheat accessions along with triticin of Chinese Spring.

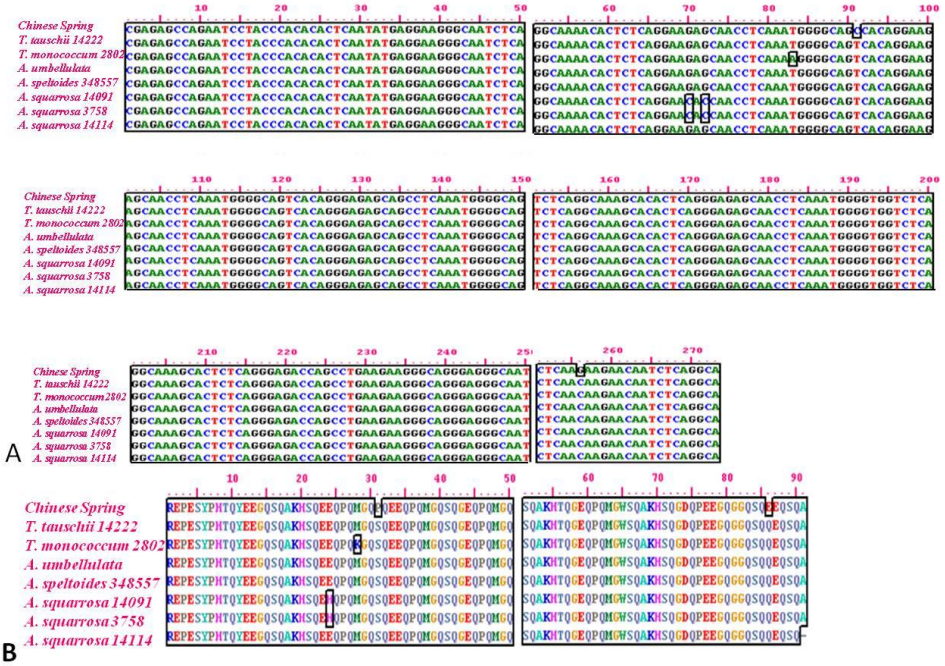

Fig 5. (A) Comparison of Chinese Spring hyper variable region sequence with seven wild wheat varieties sequence. (B) Comparison of amino acid sequence of HVR region among Chinese Spring and wild wheat accessions.

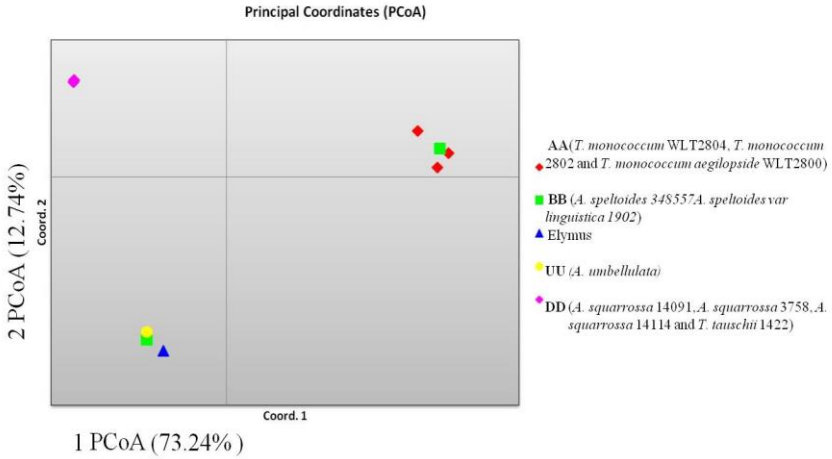

Fig 6. Principle coordinate analysis.

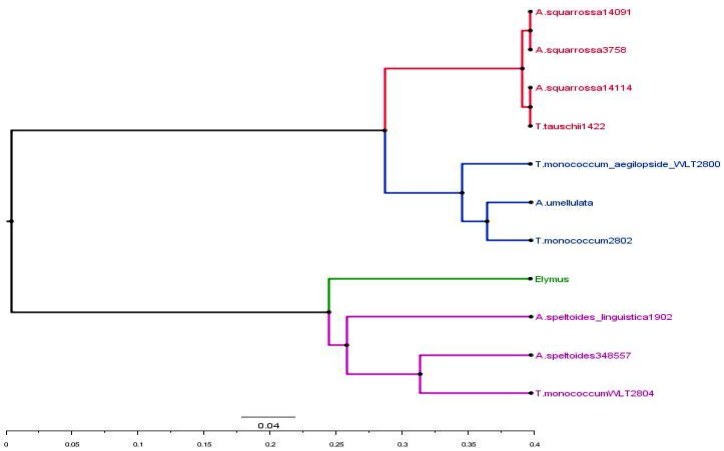

Fig 7. Dendrogram obtained by UPGMA analysis based on SNP data showing genetic relationship among the accessions.

Here, also phylogenetic analysis based on wheat SNPs revealed the closeness of U to D genome accessions hence they are falling into the same group and have similary at intronic SNP level (Fig 7). Intron1 region of D genome showed no SNPs except *T. tauschii* 4222 but intron2 showed three and Intron3 showed nine SNPs. Cluster analysis showing the total variation in the diverse wheat genome was 91.72%. First three co-ordinates explain the 73.24%, 12.74% and 5.74% of genetic variation in total eleven varieties of wheat taken on the bases of observed SNPs. The clear-cut separation of D genome was observed in both the cluster analysis and dendrogram (Fig 6, 7). In HVR, transition of C/T at the 1150th position and G/C at 1315th positions of all the seven wild wheat accessions were observed. *A. squarrosa* 14091 and *A. Squarrosa* 3758 showed G/C transition at 1129th and 1131st positions. *T. monococcum* has A instead of T at 1142nd position. Due to these transitions one accession, *T. monococcum* 2802 has an extra lysine unit in its HVR region as compared to other wild wheat accessions when compared with Chinese Spring reference sequence. Compared to exon or HVR, the introns showed greater inter specific polymorphism. In all species more SNPs were there in intron region than in HVR because of stronger selection pressure in later compared to the former. These variations may be of importance in efforts to improve seed nutritional quality by protein engineering. The naturally occurring variability in the regions implies that they do not perform critical functions in assembly of the subunits into oligomers. Therefore, it suggests that the changes engineered into the variable regions, and in particular the HVR, would be less likely to adversely affect the assembly of the subunits than changes in areas where the molecules have been highly conserved during evolution. Deletion and insertions studies in the glycinin HVR of soybean G4 subunit support this strategy because no detectable effect was reported (Dickinson, 1988) US wheat cultivars were also analyzed for SNP variation (Shiaoman Chao et al., 2009). Gliadin proteins were explored earlier for lysine content and their use as a source of wheat lysine improvement programs, but triticin has not been explored earlier for these issues (Harris, 1994). Durum wheat was also studied for lysine variation, but no significant difference was observed (Yahia Rharrabti, 2001). Triticin protein was also thought to be a renowned source, but a thorough study in different genomes has not yet been done. Here a number of different accessions of wide genomes were explored to find out triticin gene SNPs/indels which can help further study related to wheat quality improvement.

Wheat Germplasm

Twenty-one wild wheat accessions were used in the study based on their genome (A, B, D, U and SY). These samples were collected from Division of Genetics, Indian Agricultural Research Institute, New Delhi and Punjab Agricultural University, Ludhiana. The seeds were grown in the field of Genetics Division at PUSA Institute, New Delhi in October 2010 (Table 1).

DNA isolation and PCR amplification

DNA was isolated from 3-5 days old seedlings with a modified CTAB procedure (Doyle and Doyle, 1987). Primers were designed from the nucleotide sequence of Chinese Spring triticin gene of wheat. The sequence of triticin was downloaded from NCBI (GenBank:EU482413.1) and primers specific for triticin intronic and HVR were designed using the web-based tool Primer Premier 5.0 (Clarke and Gorley, 2001) corresponding to the genome sequence of triticin gene in wheat variety Chinese Spring. Three pair of primers specific to each intron region and one pair for HVR were designed (Table 2.). Amplification of intron and hypervariable region of the DNA fragments was performed using *Taq polymerase* (1 U; Qiagen, Valencia, Calif.) in 10 µl of reaction buffer (Qiagen, 1.5 mM $MgCl_2$) containing 20 ng genomic DNA, 100 mM of each dNTP and 5 picomole of each PCR primer. PCR program was run at 95 ºC for 3 minutes followed by 38 cycles of 95 ºC for 30 seconds, 55 ºC for 30 seconds and finally at 72 ºC for 1 minute. The amplified products were separated on 4% metaphor agarose gel and visualized under phospho imager system after staining with ethidium bromide. The size of the amplified fragments was estimated visually using 100 bp DNA ladder as size standard.

Sequencing of amplification products

The amplified PCR products were treated with 0.25U shrimp alkaline phosphatase and 2.5U exonuclease I at 37 ºC for 15 minutes in PCR reaction buffer (qiagen, 1.5 mM $mgCl_2$). For each sequencing reaction 100 ng (2 µl) template DNA was mixed with 4 µl DYEnamic ET mix (Amersham Biosciences), 3 µl sterile MQ water and 1 µl of forward Primer. PCR was performed for 34 cycles with denaturation of template DNA at 94 °C for 10 seconds, primer annealing at 50 °C for 10 seconds and extension at 60 °C for 2 minutes. After cleaning the PCR products, DNA samples were sequenced using Sanger's dideoxy method on automated sequencer (MegaBACE 4000). The nucleotide sequences were assembled using Phred/Phrap/Consed software tools.

In silico analysis of the triticin sequences

Sequencing was done with both forward and reverse primers and subjected to contig formation using BioEdit software. The sequences of different wild wheat accessions aligned by Clustal W and SNPs/InDels were detected. The different HVR sequences of wild wheat accessions were *in silico* analyzed for amino acid sequence with Chinese Spring reference sequence using TransSeqtool http://www.ebi.ac.uk/Tools/st/emboss_transeq/). Each sample was sequenced in triplicate with both forward and reverse gene specific primers, then further aligned using CAP contig tool of BioEdit ver 7.1 software. Beside this QC value/single peak in sequencing chromatogram was also analyzed to ensure the quality of SNP calls.

Conclusions

The results of the study showed sequence polymorphism in wild wheat accessions. Triticin Intron1 region was found to have maximum variability among different accessions followed by intron2 and intron3. Triticin has not yet been explored for its variation at sequence level. One of our major findings is the presence of an extra lysine unit in *T. monococcum* 2802 which brightens our future chances of studying such accessions where we can find more options varying in the units of lysine. Another major result is the confirmation of the triticin sequence identity of our reference variety Chinese Spring. It has been found that the triticin reference sequence is a result of D genome in Chinese Spring, which was not clear till date So further qualitative and molecular evaluation of wild wheat accessions can be utilized in breeding programs and making better nutritionally

enriched improved wheats.

Acknowledgments

We are grateful to the Wheat Quality Network Project and ICAR Network Project on Transgenic Crops for providing funding support.

References

Alvarez ML, Guelman S, Halford NG, Lustig S, Reggiardo MI, Ryabushkina N, Schewry P, Stein J and Vallejos RH (2000) Silencing of HMW glutenins in transgenic wheat expressing extra HMW subunits. Theor Appl Genet. 100:319-327.

Budak H, Kantar M, and Kurtoglu KY (2013) Drought tolerance in modern and wild wheat. Sci World J. Article ID 548246, 16 pages, doi:10.1155/2013/548246

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis models and estimation procedures. Am J Hum Genet. 19: 233–257.

Chatzav M, Peleg Z, Özturk L, Yazici A, Fahima T, Cakmak I, Saranga Y (2010) Genetic diversity for grain nutrients in wild emmer wheat: potential for wheat improvement. Ann Bot. 105: 1211–1220.

Clarke KR, Gorley NR (2001) Primer v5: User manual/tutorial. Primer-E Ltd., Plymouth, United Kingdom

Dickinson CD, *Ph.D. thesis*, Purdue University, West Lafayette, USA (1988).

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19: 11-15.

Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL (2001) Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol. 125:1342-1353.

Eggum BO, Beame RM (1983) The Nutritive Value of Seed Proteins. In: Seed Proteins Biochemistry, Genetics and Nutritive Values, Gotteschalk, W. and P.H. Muller (Eds.). Junk J.N. Publishers, The Hague.

Ferreira RR, Varisi VA, Meinhardt LW, Lea PJ and Azevedo RA (2005) Are high-lysine cereal crops still a challenge. Braz J Med Biol Res. 38: 985-994.

Gandhi HT, Vales MI, Watson CJW, Mallory-Smith CA, Mori N, et al. (2005) Chloroplast and nuclear microsatellite analysis of *Aegilops cylindrica*. Theor Appl Genet. 111: 561-572.

Garg M, Rao YS, Goyal A, Singh B (2007) Variations in seed storage protein-triticin among diploid *Triticum* and *Aegilops* species. Biotechnol. 6: 444-446.

Haider (2013) The origin of the B-genome of bread wheat (Triticum aestivum L.). Rus J Genet. 49: 263-274.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acid Symp Ser. 41:95-98.

Harris JD, Taylor GA, Blake TA, Sands DC (1994) A fluorimetric assay for determining variation of lysine in wheat gliadin proteins. Euphytica. 76:97-100.

Heo H, Ta LE (2013) Development of genome-specific primers in hexaploid. Wheat Adv Crop Sci. 3:709-715.

Ren J, Chen L, Sun D, You FM, Wang J, Peng Y, Nevo E, Beiles A, Sun D, Luo M-C and Peng J (2013) SNP-revealed genetic diversity in wild emmer wheat correlates with ecological factors. BMC Evol Biol. 13:169.

Chen L, Wang S-Q, Hu Y-G (2011) Detection of SNPs in the *VRN-A1* gene of common wheat (*Triticum aestivum* L.) by a modified Ecotilling method using agarose gel. electrophoresis. Aust J Crop Sci. 5(3): 321-329.

Patnaik D, Khurana P (2001) Wheat biotechnology: A minireview. Electron J Biotechnol. 4(2): 94-100.

Ramakrishna W, Dubcovsky J, Park YJ, Busso C, Embereton J, SanMiguel P, Bennetzen JL (2002) Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. Genetics. 162:1389-1400.

Rharrabti Y, Villegas D, García del Moral LF, Aparicio N, Elhani S, Royo C (2001) Environmental and genetic determination of protein content and grain yield in durum wheat under Mediterranean conditions. Plant Breed. 120:381-388.

Shailja K, Rathore M, Puri N, Yadav D, Singh NK (2002) PCR amplification of hypervariable region of wheat triticin gene. J Cereal Sci. 35:129-134.

Chao S, Zhang W, Akhunov E, Sherman J, Ma Y, Luo M-C, Dubcovsky J (2009) Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. Mole Breed. 23:23-33.

Singh BK, Singh A, Gaikwad K, Singh NK (2009) Cloning and characterization of full-length triticin cDNA and genes from wheat varieties K-68 and Chinese Spring. J Plant Biochem Biotechn. 18 (1):21-28.

Singh NK, Donovan GR, Carpenter HC, Skerritt JH, Langridge P (1993) Isolation and characterization of wheat triticin cDNA revealing a unique lysine rich repetitive domain. Plant Mole Biol. 22:227-237.

Singh NK, Shepherd KW, Langridge P, Gruen LC, Skerritt JH and Wrigley CW (1988) Identification of legumin like protein in wheat. Plant Mol Biol. 11: 633-639.

Syvanen A-C (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet. 2:930-942.

Thompson DJ, Higgins DG, 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(22): 4673-4680.

Wang SL, Qi LL, Chen PD, Liu DJ, Friebe B, Gill BS (1999) Molecular cytogenetic identification of wheat-Elymus tsukushiense introgression lines. Euphytica. 107: 217-224.