Journal of Theoretical Biology 376 (2015) 1–7

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Molecular characterization of the puroindoline-a and b alleles in synthetic hexaploid wheats and in silico functional and structural insights into *Pina-D1*

Iftikhar Ali [a], Zainab Sardar [b], Awais Rasheed [a], Tariq Mahmood [a,*]

[a] Department of Plant Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan
[b] Department of Botany, Government Jahanzeb Postgraduate College Saidu Sharif, Swat, Pakistan

HIGHLIGHTS

- Grain hardness is determined by tightly linked Puroindoline genes, *Pina* and *Pinb*.
- Three new *Pina* alleles were found and were designated as *Pina-D1w*, *Pina-D1x* and *Pina-D1y*.
- Most of the structural features of *Pina* were found conserved.
- In silico structural and functional characteristics of *Pina-D1* were highlighted.
- PINA have binding capacity with small parts of prolamins causing celiac disease of human.

ARTICLE INFO

Article history:
Received 29 September 2014
Received in revised form
22 March 2015
Accepted 1 April 2015
Available online 10 April 2015

Keywords:
Puroindoline a (*Pina-D1*) gene
Allelic diversity
Protein modeling
Grain hardness
Synthetic hexaploid wheat

ABSTRACT

Kernel hardness determined by two tightly linked Puroindoline genes, *Pina-D1* and *Pinb-D1*, located on chromosome 5DS define commercially important characteristics, uses, major grades and export markets of wheat. This study was conducted to characterize *Pina-D1* and *Pinb-D1* alleles, in fifteen synthetic hexaploid wheats (SHWs) and its relation with grain hardness. Additionally, in silico functional analyses of puroindoline-a protein was conducted for better understanding of their putative importance in grain quality. Six different *Pina-D1* alleles were identified in the SHWs, of which three i.e. *Pina-D1a*, *Pina-D1c* and *Pina-D1d* were already known whereas the other three had new sequence polymorphisms and were designated as *Pina-D1w*, *Pina-D1x* and *Pina-D1y*. Three different *Pinb-D1* alleles were identified which have been reported earlier and no novel sequence polymorphism was detected. It was concluded that despite some primary, secondary and 3D structure variations, ligand binding sites and disulfide bonds discrepancies, the main features of PINA, i.e. the tryptophan-rich domain, the cysteine backbone, the signal peptide and basic identity of the proteins were all conserved. In silico analysis showed that puroindolines having binding capacity with small parts of prolamins causing celiac disease of human, however their potential role is not obvious. Conclusively, the new *Pina-D1* alleles with modest effect on grain hardness, and insight into their functional and structural characteristics are important findings and their putative role in celiac disease require further studies to validate.

1. Introduction

Wheat (*Triticum aestivum* L.) is a self-pollinated, annual, most extensively cultivated and consumed crop worldwide. Synthetic hexaploid wheats (SHWs) are the result of artificial intercrossing between *Triticum turgidum* and *Aegilops tauschii* accessions, the evolutionary ancestors of common wheat (Mujeeb-Kazi et al., 1996). SHWs can act as a vehicle for introducing specific characters from the various *A. tauschii* accessions into bread wheat. The hybridization events that formed bread wheat are thought to be limited, thus the genetic diversity within the SHWs possess novel alleles and genes for abiotic, biotic stress tolerances, and grain quality traits, not currently represented within the bread wheat gene pool (Ogbonnaya et al., 2013).

Wheat grain quality is the single most important trait that determines the market value and is operated across the world based on the features of kernel texture. Soft wheat have softer endosperm texture that ruptures easily, so need less energy to mill, produce smaller particles, and small amount of starch is damaged

* Corresponding author.
E-mail addresses: tmahmood.qau@gmail.com, tmahmood@qau.edu.pk (T. Mahmood).

http://dx.doi.org/10.1016/j.jtbi.2015.04.001
0022-5193/

after milling as compared to hard wheat (Symes, 1961). In the past 20 years, the genetic basis for grain hardness in wheat has been resolved, and shown that it is mostly measured by the degree of association between the protein matrix and the starch granules that is controlled by the complex friabilin proteins (Beecher et al., 2002). Further study revealed that friabilin were primarily composed of *Pins* (*Pina-D1* and *Pinb-D1*), and genes are located on the *Ha* locus on chromosome 5DS (Morris et al., 1994).

Pina-D1 and *Pinb-D1* genes were cloned and have 70.2% sequence similarity in the intronless coding regions consisted of 447 bp (Gautier et al., 1994). Soft bread wheats have wild type *Pina-D1a* and *Pinb-D1a* alleles and mutation in either of these alleles result in hard texture with the alleles being designated *Pina-D1b* and/or *Pinb-D1b*. Geng et al. (2013) demonstrated that the *Pina-D1b/Pinb-D1a* genotype is 10 single kernel characterization system (SKCS) hardness index units harder than the *Pina-D1a/Pinb-D1b* genotype, and the hardness index ranking was b/b > b/a > a/b > a/a. However *Pinb-D1b* is known to increase flour yield, lower ash, increase loaf volume and improve crumb grain score (Hogg et al., 2005). Gedye et al. (2004) suggested that several *Pin-D1* genotypes are available in *A. tauschii* which give a wide range of extra soft to soft grain textures.

The proteins encoded by both genes, PINA and PINB, are small (~13 kDa), highly basic (pI 10–11), cysteine (Cys) and tryptophan-rich seed-specific isoforms, composed of 148 amino acids that display an amphiphilic TRD (5 Trp in PINA and 3 in PINB) with adjacent arginine (Arg), belong to a widespread 2S superfamily of endosperm seed proteins (Gautier et al., 1994; Shewry et al., 2004). Grain texture of hard wheat is affected by the amount of PINs in each allele. *Pina-D1* and *Pinb-D1* alleles have diverse amount of PINs; *Pina-D1a*/*Pinb-D1b* haplotype has normal quantity of PINA but reduced quantity of PINB, which is completely absent from *Pinb-D1c* and *Pinb-D1p*. *Pina-D1b* lacked PINA, whereas *Pina-D1k* lacked both of PINA and PINB. So the expression patterns of PINA and PINB differ among *Pin* alleles (Ikeda et al., 2010). However, large variations in grain hardness are often noticed with the same allele which indicated the presence of other minor genes underpin grain hardness (Chen et al., 2007). To date, 22 *Pina-D1* and 25 *Pinb-D1* alleles designated as *Pina-D1a-t* and *Pinb-D1a-w, aa, bb* respectively have been reported from different common wheat and other related diploid and hexaploid wheat varieties. Present study was conducted to characterize *Pina-D1*and *Pinb-D1* alleles in the 15 SHWs, its relation with grain hardness, and in silico structural and functional analysis of its coded proteins.

2. Materials and methods

2.1. Synthetic wheat germplasm

The seeds of fifteen SHW accessions (Supplementary data; Table 1) were sown in Jiffy® under controlled conditions of light, temperature and humidity in a growth room during 2012–13. Total genomic DNA of all samples was extracted from young leaves of 2–3 weeks old seedling using phenol / chloroform method. The complete coding region of *Pina-D1* was amplified by a pair of oligonucleotide PCR primers (Forward primer: 5′-ATGAAGGCCCTCTTCCTCA-3′ and Reverse primer: 5′-TCACCAGTAATAGCCAATAGTG-3′), coding sequence of *Pinb-D1* was amplified by a pair of primers (Forward primer: 5′-ATGAAGACCTTATTCCTCCTA-3′ and Reverse primer: 5′-TCACCAGTAATAGCCACTAGGGAA-3′) already designed by Gautier et al. (1994). PCR amplifications of *Pina-D1* and *Pinb-D1* genes were carried out in a final volume of 25 μl reaction mixture containing 10 mM of PCR buffer (20 mM Tris–HCl pH 8.4, 20 mM NH_2SO_4, 1.5 mM $MgCl_2$), 200 μM of each dNTPs, 8 pmol of each primer and 30 ng of genomic DNA and 1 U of *Taq* polymerase (Tiangen Biotech, Beijing, China). Final volume was reached to 25 μl by adding nuclease free double distilled PCR water. The PCR cycling parameters consisted of an initial denaturation of 95 °C for 3 min, followed by 35 cycles of 95 °C for 1 min, 55 °C for 1 min, 72 °C for 2 min, and a final extension at 72 °C for 10 min. PCR amplified product was purified by JetQuick™ PCR purification kit (Genome GmbH, Germnay), visualized with UV light (Supplementary data; Fig. 1b) using 1.5% agarose gel, and were sequenced commercially from Macrogen Inc. (Korea).

2.2. Measurement of grain hardness

One hundred randomly selected seeds of each SHW accession were ground through UDY cyclone mill (Seedburo Equipment Co., IL) fitted with 0.5 mm sieve for whole wheat flour production. The mill was carefully cleaned after each sample to avoid mixing of the other samples. Near-Infrared Reflectance (NIR) spectroscopy was used to measure grain hardness, by running each SHW accession through NIR Analyzer Inframatic 8620 (Perten Instruments, Inc., IL, USA) according to the procedure described in AACC method no. 39-70 A. (AACC, 1999) Average and standard deviation was obtained from these data to classify the varieties into its texture types.

2.3. Computational analysis of the sequence data

Mutation surveyor (http://www.softgenetics.com/mutationsurveyor.html) was used to determine mutations from Sanger sequencing peaks of *Pina-D1 and Pinb-D1* gene. Signal peptide, functional peptide and cleavage site of PINA proteins were determined by online tool 'Signal-BLAST' (http://sigpep.services.came.sbg.ac.at/signalblast.html/). 3D atomic models of PINA were determined by I-TASSER (http//zhanglab.ccmb.med.umich.edu/i-TASSER) (Zhang, 2010), and then visualized and analyzed in atomic, ionic and crystal forms by Jmol (http://jmol.sourceforge.net/). Structure based biological function annotations were predicted by COFACTOR (http://zhanglab.ccmb.med.umich.Edu/COFACTOR/). InterProScan (version 4.8) was used to determine functional sites or domains of PINA (http://www.ebi.ac.uk/Tools/pfa/iprscan/).

For PINA-protein interaction, protein sequence (GenBank accession P18573) of wild type *Pina-D1* allele was selected for docking. PINA-prolamin docking was analyzed by HADDOCK web-server freely available (http://haddock.chem.uu.nl/). Active residues (interacting residue) were identified with the help of another tool "CPORT" (http://haddock.science.uu.nl/services/CPORT/). HADDOCK Webserver evaluates models on the basis of HADDOCK score. The score is a combination of buried surface area, Desolvation, electrostatic, Vander Waals and restraint violation energies (De Vries et al., 2010).

3. Results

3.1. Sequence polymorphism in Pina-D1 gene

Six SHW accessions (SH-326, SH-381, SH-395, SH-500, SH-505 and SH-675) had wild type *Pina-D1* allele and carry no mutation in coding sequence. All other SHW accessions have mutations in coding sequence of *Pina-D1* locus, of which the most common was single nucleotide polymorphism (SNP) G/A at 257 bp position identified in SH-302, SH-403, SH-424, SH-618, SH-638, SH-649, SH-677, SH-678 and SH-852 accessions. The *Pina-D1* allele from SH-618, SH-638, SH-677, SH-678 and SH-852 had one non-synonymous SNP G/A at nucleotide position 257, which changed the codon CGG (Arg) to CAG (Gln) at amino acid position 58. In SH-649, two SNPs were found in *Pina-D1* sequence i.e. G/A at nucleotide position 57 and 257, which resulted in the synonymous