Journal of Cereal Science 48 (2008) 722–728

Contents lists available at ScienceDirect

Journal of Cereal Science

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

Identification and genetic mapping of variant forms of puroindoline b expressed in developing wheat grain

Mark Wilkinson [a], Yongfang Wan [a], Paola Tosi [a], Michelle Leverington [b], John Snape [b], Rowan A.C. Mitchell [a], Peter R. Shewry [a,*]

[a] Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK
[b] John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

ARTICLE INFO

Article history:
Received 2 January 2008
Received in revised form 11 March 2008
Accepted 18 March 2008

Keywords:
Wheat
Grain hardness
Puroindolines
Grain development
Genetic mapping

ABSTRACT

Transcripts encoding three novel variant forms of puroindoline b have been identified in developing seeds of wheat. These show 57–60% sequence identity with the wild type form of Pin b but all lack one of the three tryptophan residues present in the "tryptophan loop" region of the wild type protein. Counts of ESTs and array analysis indicate that the transcripts encoding variant forms of Pin b are about an order of magnitude less abundant than those encoding wild type Pin b while array analysis also shows that expression of the variant form 1 declines more rapidly than that of the wild type form during the later stages of grain development. The gene(s) encoding variant form 1, named *Pinb-A2*, were mapped to the long arm of chromosome 7A of bread wheat where they show linkage to novel QTLs for hardness which have been identified in two doubled haploid populations derived from crosses between hard parental cultivars (Shamrock × Shango, Malacca × Charger).

1. Introduction

Puroindolines (Pins) are a group of proteins which are restricted to seeds of wheat and related cereal species. They belong to an extensive superfamily of plant proteins which includes a number of other cereal seed proteins but are characterised by the presence of a short tryptophan-rich sequence (Douliez et al., 2000; Shewry et al., 2004). Two types of Pins are present in wheat, which differ mainly in the presence of either five (Pin a) or three (Pin b) tryptophan residues (Blochet et al., 1993). Although Pins were discovered as recently as 1993 they have since become one of the most intensively studied groups of cereal seed proteins. This is because of their role in determining grain texture, a property of fundamental importance in determining the milling and baking performance of wheat (Pomeranz and Williams, 1990).

A range of studies have been carried out including extensive surveys of allelic variation in Pins in lines differing in grain texture and analyses of Pin alleles in the progeny of crosses between hard and soft genotypes (reviewed by Bhave and Morris, 2008a,b; Morris, 2002). These studies have shown that the expression of genes encoding the wild type forms of Pin a and Pin b results in soft texture while hardness is associated with two types of mutations. These are either "null" mutations which result in the silencing of the genes encoding Pin a and/or Pin b or mutations which affect the amino acid sequence of the Pin b protein.

The structural genes encoding Pin a and Pin b map to the major *Hardness* (*Ha*) locus, which is located on chromosome 5D of bread wheat (Law et al., 1978), and nucleotide sequencing has shown that this locus comprises three genes within a region of about 82,000 bp (Chantret et al., 2005). These are the *Pina* and *Pinb* structural genes and the *Gsp-1* gene encoding a protein called "grain softness protein" (GSP) which is structurally related to the Pins.

Genes related to *Pina* and *Pinb* are present in diploid species with genomes related to the A and B genomes of bread wheat but they appear to have been lost by independent deletions from the A and B genomes of bread wheat and tetraploid *Triticum turgidum* which includes cultivated durum (pasta) wheat. As a result durum wheat is harder in texture than even the hardest bread wheat genotypes. The fact that GSP genes are present, and presumably expressed, in durum wheat indicates that it does not play a major

Abbreviations: CE, capillary electrophoresis; dbEST, GenBank Expressed Sequence Tag Database; DHLs, doubled haploid lines; ESTs, expressed sequence tags; GSP, grain softness protein; MDE, mutation detection enhancement matrix; MOPS, 3-[N-morpholino]propanesulfonic acid; NIR, near infra-red; PCR, polymerase chain reaction; QTL, quantitative trait locus; SKCS, single kernel characterisation system; SSCP, single strand conformation polymorphism; SSR, simple sequence repeat; TEMED, tetramethylethylene diamine; UV, ultraviolet.

* Corresponding author. Tel.: +44 1582 763133; fax: +44 1582 763010.
E-mail address: peter.shewry@bbsrc.ac.uk (P.R. Shewry).

0733-5210/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jcs.2008.03.007

role in determining grain texture (Chantret et al., 2005) and this is supported by the analysis of genetic stocks of cv. Chinese Spring differing in GSP content (Tranquilli et al., 2002).

Nevertheless genetic studies show that the *Ha* locus does not account for the full range of variation in grain texture in bread wheat, and minor quantitative trait loci (QTLs) have been mapped on several chromosomes (Galande et al., 2001; Law et al., 1978; Sourdille et al., 1996; Turner et al., 2004).

We are therefore interested in identifying novel components which may contribute to variation in grain texture of wheat and report the characterisation of transcripts encoding novel variant forms of Pin b which are expressed in the developing wheat grain. The biological role of these proteins has not been established but genetic analysis shows that they map to chromosome 7A which is also the location of minor hardness QTLs identified in two crosses between cultivars with hard grain texture.

2. Materials and methods

2.1. PCR amplification and sequence analysis

Total RNA for RT-PCR was isolated from developing whole grains of bread wheat cv. Hereward (at 14 days after anthesis) and durum wheat cv. Ofanto (at 21 days after anthesis) following the method of Chang et al. (1993). The integrity of RNA was checked on 1.5% MOPs gel and an Agilent Bioanalyser 2000.

For cDNA synthesis, 5 μg of total RNA was digested with RNase-free DNase (from Promega or Ambion) to remove genomic DNA, according to the manufacturer's instructions. The DNase-treated RNA was reverse transcribed into cDNA using oligo-d$(T)_{23}$ primer (Sigma) and SuperScript™ III reverse transcriptase (Invitrogen). Twenty-five nanograms of cDNA was used for PCR amplification of Pin b variants.

Genomic DNA was extracted from leaf material of seedlings using a Promega Wizard kit according to the manufacturer's instructions. The primers and conditions for PCR were the same as those used for cDNA amplification.

Full-length Pin variant forms 1 and 2 were amplified with gene-specific primers. The forward primers used to amplify Pin b variants 1 and 2 were: Pinb-VF-5′-ATGAAGACCTTATTCCTCCTAGCTC-3′. The reverse primer for Pin b variant 1 was a degenerate primer Pinb-VR1-5′-TCAGTAGTAATAGCCATTAGTAKGGACG-3′ while the reverse primer for Pin b variant 2 was also degenerate, Pinb-VR2-5′-TCACTAGTAATAGCCATTAKTAGCGACA-3′, where K is G + T.

The reactions were performed in 25 μl containing approximately 200 ng of cDNA, 0.5 μM of each primer, 200 μM of each dNTP, 1× Phusion DNA polymerase reaction buffer and 0.03 U/ml of Phusion DNA polymerase (Finnzymes Oy, Keilaranta 16A, 02150 Espoo, Finland). The amplification conditions used for cv. Hereward were 40 cycles at 96 °C for 15 min (Qiagen Hotstar Taq DNA polymerase), 30 s at 96 °C, 1 min at 50 °C, 1 min at 72 °C and a final extension of 7 min at 72 °C. The conditions for cv. Ofanto were 98 °C for 3 min, 35 cycles at 98 °C for 17.5 min, 30 s at 60 °C, 90 s at 72 °C and a final extension of 10 min at 72 °C.

Five microlitres of the PCR products were analysed on 1.0% (w/v) agarose gels, stained with ethidium bromide and visualised by UV light. PCR products were purified using a Wizard® SV Gel and PCR Clean-Up System (Promega). Fragments were then ligated into pGem®-T Easy Vector System (Promega) and clones obtained using the Wizard® Plus SV Minipreps DNA purification system (Promega). All of the above procedures were carried out according to the manufacturer's instructions. Sequencing reactions were performed with the BigDye Terminator Version 3.1 Cycle Sequencing Kit (ABI) and all reactions were analysed at Geneservice (Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU). Annotations of sequences were carried out using the Sequencher 4.1 software package (MAC).

Phylogenetic trees were derived from multiple alignments of protein sequences with the MUSCLE algorithm (Edgar, 2004), removal of gapped columns followed by phylogeny analysis in the phyml package (Guindon et al., 2005) using the Whelan and Goldman (2001) model which was found to give the highest likelihood of the models tested. An initial run optimised the gamma and invariant proportion parameters to 3.09 and 0.08, respectively. These were then held constant for 1000 runs for bootstrap non-parametric analysis.

2.2. Array analysis

The wheat Affymetrix Genechip® array data sets on developing whole grain of hexaplod wheat cv. Hereward described by Wan et al. (2008) were normalised using gcRMA (Wu et al., 2004) and analysed using GeneSpring GX 7.0. These data are deposited in the ArrayExpress database (accession E-MEXP-1193).

2.3. Mapping and linkage analysis

To map the new *Pin* locus, named *Pinb-2*, in bread wheat, two alternate but overlapping sets of primers were designed from the 3′ untranslated region. Both sets of primers amplified polymorphic products on the parental varieties Spark and Rialto at an annealing temperature of 52 °C, and showed the same scores on an initial test of 30 doubled haploid lines. The Pinb-2-1 forward primer was 5′-ACCTTATTCCTCCTAGCTCTC-3′ and the reverse primer 5′-TAGTTGCTGGCAACACTGGTC-3′. The Pinb-2-2 forward primer was 5′-CGGTGCTTGGCTGTGAAG-3′ and the reverse primer 5′-CAC-TAGTAATAGCCATTATTAGCGAC-3′.

Single strand conformation polymorphism (SSCP) analysis using the mutation detection enhancement (MDE™) matrix was used to detect sequence polymorphisms in the PCR amplicons. Samples were denatured at 95 °C for 3 min, plunged into an ice water bath, and loaded onto standard sequencing size gels (30 cm × 40 cm × 0.4 mm) containing 12.5 ml MDE gel solution (Cambrex Bio Science), 2 ml 20× TTE buffer (National Diagnostics), 27 ml water, and 9 ml 50% glycerol solution, polymerised by the addition of 300 μl 10% (w/v) ammonium persulphate and 30 μl TEMED. Electrophoresis was carried out for 17 h at a constant power of 4.5 W at 5 °C, and gels were stained with silver (Bassam et al., 1991).

For mapping the intra-chromosomal location of the new *Pinb-2* locus, the locus was genotyped on three recombinant doubled haploid populations: Spark × Rialto, Shamrock × Shango and Malacca × Charger. These genotyping data were combined with already available extensive genotype files based on SSR and DArT markers on these populations. Maps were then re-calculated using the mapping software JoinMap Vs 3.0 to position the Pin locus. The LOD threshold selected was between 4 and 6, depending on the linkage group. The distances for recombination were calculated according to the Kosambi mapping function.

The grain hardness of the whole populations of Shamrock × Shango and Malacca × Charger recombinant doubled haploid lines was measured on milled grain by NIR using a Bran and Lübbe NIR-Spectrometer Infralyser 2000 and calibrations validated by comparison with measurements made with the Perten Single Kernel Characterisation System (SKCS) (Turner et al., 2004). These data were obtained from pooled seed of each doubled haploid line grown in two separate randomised and replicated experiments (three replicates per site) at two sites in Eastern England in 2005. The QTL analysis for hardness was carried out on the phenotypic data obtained from the NIR measurements and analysed using QTL Café software. The hardness of the parental lines from the same field plots was also measured using the Perten SKCS.