Theor Appl Genet (2005) 110: 1259–1267
DOI 10.1007/s00122-005-1962-3

ORIGINAL PAPER

Bryan Clarke · Sadequr Rahman

A microarray analysis of wheat grain hardness

Received: 12 July 2004 / Accepted: 14 February 2005 / Published online: 2 April 2005

Abstract Grain hardness is an important quality characteristic of wheat grain, and considerable research effort has focused on characterising the genetic and biochemical basis underlying the hardness phenotype. Previous research has shown that the predominant difference between hard and soft seeds is linked to the puroindoline (PIN) proteins. In this study the near-isogenic lines of Heron and Falcon, which differ only in the grain hardness character, were compared using a cDNA microarray consisting of approximately 5,000 unique cDNA clones that were isolated from wheat and barley endosperm tissue. Our analysis showed that major differences in gene expression were evident for puroindoline-a (*Pina*), with a minor but not consistent change in the expression of puroindoline-b (*Pinb*). These observations were confirmed using a 16,000 unique cDNA microarray in a comparison of hard wheats with either the *Pina* null or *Pinb* mutation.

Introduction

Grain hardness in hexaploid wheat (*Triticum aestivum* L.) is a major determinant of the end-use properties for the grain. The hardness of the grain determines the starch damage during milling, and this in turn largely determines the optimum water requirement during the baking process; hard and soft wheats are used for different types of products (Morris 2002; Turnbull and Rahman 2002). Hardness is usually measured using the single kernel characterisation system (SKCS), which measures the force required to crush grain between a toothed rotor and a progressively narrowing crescent gap. However, more indirect measurements have also been used, such as the particle size index, which is a measure of the particle size distribution after grinding the grain (Symes 1961) and the electron microscopy of cross-sections of the grain, which was used to differentiate soft and hard wheats (Bechtel and Wilson 1997; Turnbull et al. 2002). Differences in hardness can be detected in the developing grain after air-drying, demonstrating that it is a property that is observable early in development (Bechtel et al. 1996; Turnbull et al. 2002).

The genetic basis of hardness has been investigated for some time. Symes (1965), using the particle size index as a measure for grain hardness, demonstrated the segregation of a major gene for this trait between the wheat cultivars Heron and Falcon. Results obtained in the same study from seven other crosses between hard and soft wheat also indicated that a single gene was responsible for most of the difference in hardness. The gene influencing hardness, named *Ha*, was subsequently mapped to the short arm of chromosome 5D (Mattern et al. 1973). Although the locus is referred to as "Hardness" (*Ha*), "Softness" is the trait actually conferred by this locus.

The biochemical basis of grain hardness is less well understood. An important advance came with the observation of Greenwell and Schofield (1986) that the presence of a 15-kDa protein on water-washed wheat starch was correlated with grain softness for 150 different wheats. These authors found that there was more of the 15-kDa protein associated with water-washed starch from soft wheat, less with hard and none with durum (*T. turgidum* spp. durum) (very hard) wheats. The 15-kDa protein was named "friabilin", indicating the fact that soft wheats are more easily crumbled than hard (Greenwell and Schofield 1989). Jolly et al. (1993) showed that the 15-kDa "protein" was in fact composed

Communicated by J.W. Snape

B. Clarke (✉) · S. Rahman
CSIRO Division of Plant Industry, Clunies Ross St,
Canberra, ACT 2601, Australia
E-mail: bryan.clarke@csiro.au
Tel.: +61-2-62465054
Fax: +61-2-62465000

B. Clarke
NSWCAG Division of Plant Industry,
BPO Box 1600, Canberra, ACT 2601, Australia

of a number of polypeptides. They raised antibodies against this 15-kDa polypeptide mixture and showed that it occurred in soft and hard wheat endosperm in approximately equal amounts. However, during the aqueous isolation of wheat starch as little as 1% of the original friabilin partitions with the starch of soft wheats and much less with hard ones. Greenblatt et al. (1995) devised a method for the selective extraction of the components of the 15-kDa friabilin protein and also observed that hard wheats have less friabilin attached to the washed starch granules than soft wheats. Blochet et al. (1993) used the N-terminal amino acid sequence of Triton-X-extractable proteins from wheat endosperm and obtained two major sequences, which they called puroindoline-a (PINA) and puroindoline-b (PINB). These sequences were found to correspond to those determined by Jolly et al. (1993) for the major components of the 15-kDa protein. A third protein, grain softness protein (GSP-1), was also identified as a component of the 15-kDa fraction by Rahman et al. (1994) using immunological screening of cDNA expression libraries. More recently, Oda and Schofield (1997), using two-dimensional electrophoresis, identified the 0.19 and the 0.53 alpha-amylase inhibitors in addition to the PINA and PINB polypeptides in friabilin. Thus, the 15-kDa protein complex identified by Greenwell and Schofield (1986) consists of at least five components: PINA, PINB, GSP-1 and the alpha-amylase inhibitors 0.19 and 0.53.

The genes *Pina*, *Pinb* and *Gsp-1* have been shown to be tightly linked to the *Ha* locus on chromosome 5D (Giroux and Morris 1997, 1998; Giroux et al. 2000; Jolly et al. 1996; Sourdille et al. 1996). Detailed analysis of a bacterial artificial chromosome (BAC) covering that region showed that the genes were in the order *Pinb*, *Pina* and *Gsp-1* within a 60-kb fragment from the D genome donor of wheat (*Aegilops tauschii*) (Turnbull et al. 2003). Interestingly, this same arrangement has been observed in *T. monococcum* (Chantret et al. 2004; Tranquilli et al. 1999), which is more closely related to the A-genome donor of wheat. A significant advance in the understanding of the nature of the *Ha* locus came with the demonstration that in the hard wheats examined there was a mutation in either *Pina* or *Pinb* (but not in *Gsp-1*), and a direct functional role for the *Pin* genes has been shown by transformation experiments. Krishnamurthy and Giroux (2001) and Beecher et al. (2002) reported that rice and wheat transformed with the *Pin* genes were softer than untransformed controls. However, the biochemical basis of action of the PIN proteins is not known. In addition, other loci separate from the *Ha* locus may also be involved in modifying grain hardness. For example, chromosomes 1A and 6D also affect grain hardness (Perretant et al. 2000).

The introduction of cDNA microarray technology has enabled the simultaneous comparison of thousands of genes. Using this system, the transcription profile and differential gene expression of two otherwise similar cultivars that vary in a particular trait can be examined. As a source of genes, the wheat and barley endosperm cDNA libraries (Ali et al. 2000; Clarke et al. 2000) provide an extensive range of genes for analysis. Also, the near-isogenic lines (NIL) of Heron and Falcon (Symes 1969), which vary in the presence or absence of grain hardness, are valuable material for studying the genetic basis of the hardiness trait. The Falcon parent used in the generation of these stocks is known to be missing the *Pina* gene (Giroux and Morris 1998). However, differences in the expression of other components cannot be ruled out, especially as there are apparent differences in the timing of the appearance of grain hardness and the onset of *Pin* expression.

Given the complex mixture that comprises friabilin, and the indication that other chromosomal locations may be involved in the hardness phenotype, a genomic approach to the analysis of grain hardness may give new insights into this important characteristic. In this study we use cDNA microarray technology to investigate the transcriptional profiles and differential gene expression between hard and soft seed in both the Heron and Falcon backgrounds. The expression profiles are evaluated for three stages of endosperm development. A comparison is then made among transcription profiles from a range of hard and soft wheat cultivars at an early stage of endosperm development.

Materials and methods

Plant materials

The two sets of soft and hard NILs of *Triticum aestivum* in the Falcon and Heron backgrounds (Symes 1961, 1969) were used. To generate these NILs, the hard and soft phenotypes were backcrossed seven times using either Falcon or Heron as the recurrent backcross parent, and at this stage it is expected that less than 0.5% of the donor genome should remain. The seed planted from this backcrossed material was Heron hard AUS90277 and soft AUS90286, and Falcon hard AUS 90264 and soft AUS 90252. In addition, *T. aestivum* cv. Heron (AUS 322), Falcon (AUS 206), Corrigin (AUS 24850), Cook (AUS 20275), Kulin (AUS 23163), Janz (AUS 24794), Hartog (AUS 21533), Egret (AUS 16037) and Sunstar (AUS 22177) were used.

Establishment of non-redundant cDNA clones for microarray

Clones were selected from three cDNA libraries for the construction of the microarray. These libraries were a non-normalised endosperm library from *T. aestivum* cv. Wyuna (Clarke et al. 2000) and two normalised endosperm libraries; one from *T. aestivum* cv. Hartog and the other from *Hordeum vulgare* cv. Himalaya (Ali et al. 2000). The isolated cDNA clones were sequenced and annotated using BLAST as detailed in Clarke et al. (2000).

To establish the non-redundant gene set, the clipped sequences were assembled with GCPHRAP (P. Green, University of Washington Genome Centre, unpublished; see http://www.genome.washington.edu/UWGC/analysistools/phrap.htm), and assembled sequence alignments were viewed with the GAP4 programme (Staden et al. 1998). The cDNA containing the longest transcript, identified from the sequence alignments, was selected for the array. To further reduce redundancy, the BLAST output from the selected sequence was used to remove those sequences with the same BLAST annotation, but only if the similarity between the sequences was 80% or greater. Following this analysis, a total of 5,047 unique clones were obtained.

Construction of endosperm array

The non-redundant set of cDNA clones were grown overnight in 250 ml TYPGN medium [20 g tryptone, 19 g yeast extract, 10 ml 80% (v/v) glycerol, 5 g Na_2HPO_4, 10 g KNO_3 per litre] supplemented with 100 μg/ml ampicillin, in 96-well microtiter plates. These plates were replicated in 150 μl of fresh medium and grown overnight. The replicate plate was stored at 80°C after the addition of 100 μl 80% (v/v) glycerol. Plasmid preparations were made from cultures following standard alkaline lysis procedures, and the inserts were amplified by PCR using the M13/pUC forward and reverse sequencing primers (New England Biolabs, Beverly, Mass.). For quantitative and qualitative evaluation of the amplification procedure, 1 μl of the PCR reaction was run on a 2% (w/v) agarose gel in Tris-borate buffer, pH 8, and the image recorded. The remainder of the PCR products were ethanol-precipitated, resuspended in 6 μl 50% (w/v) dimethyl sulfoxide and transferred from the 96-well plates to 384-well microtiter plates for printing. The PCR products were printed onto CMT-GAPS-coated microscope slides (Corning, USA) using an Omnigrid Microarrayer (Genemachines, San Carlos, Calif.) with ChipMaker2 quill pins (TeleChem, Sunnyvale, Calif.). Prior to hybridisation the slides were baked at 80°C for 2 h.

The New South Wales centre for agricultural genomics (NSWCAG) array

This array was constructed following the same procedure as that used for the construction of the endosperm array. The array consisted of the following cDNA clones: (1) approximately 6,000 expressed sequence tag (EST) clones which mapped to the wheat genome (kindly supplied by O. Anderson, Western Region Research Centre, USDA, Albany, Calif., USA); (2) a unigene set of 9,000 EST clones from Bristol University (see http://www.cerealsdb.uk.net/wheat.htm for details of the libraries used) that were kindly supplied by Keith J. Edwards, School of Biological Sciences, University of Bristol. Bristol, UK; (3) a collection of 400 ESTs from drought-stressed wheat seedlings (kindly supplied by Heather Way and Xue Gang-Ping, CSIRO PI, St. Lucia, Brisbane, Australia); (4) the 5,047 mid-development-stage wheat and barley endosperm clones discussed above.

This new array contained approximately 22,000 clones of which 16,000 were unique (unique = less than 80% sequence similarity at the nucleotide level).

RNA isolation labelling and slide hybridisation

For both the Heron and Falcon NILs, whole wheat heads were harvested at 6, 15 and 25 days post-anthesis (DPA) from glasshouse-grown material and stored at 80°C. For the other wheat cultivars analysed, heads at approximately 9–10 DPA were used. Eight to ten developing caryopses at the same age and morphological stage of development were selected from a head to represent a sample. Seed collected from a different head represented a replicate sample. Total RNA was isolated from the whole seed following the method of Higgins et al. (1976) with the following modification to remove starch. Following the lithium chloride precipitation, the RNA pellet was resuspended in 250 μl water. Then 3.5 μl 3 *M* sodium acetate (pH 5.3) and 125 μl ethanol were added and the sample centrifuged for 10 min at 4°C. The supernatant was transferred to a fresh tube and the RNA precipitated by the addition of 21.5 μl 3 *M* sodium acetate (pH 5.3) and 375 μl ethanol. The recovery of RNA varied from 0.06 μg RNA to 0.5 μg RNA per milligram tissue, and this variation was dependent on the developmental stage of the seeds used for extraction of the RNA.

For the labelling procedure, 50 μg of total RNA was used for both the Cy3 and Cy5 dyes (Amersham Pharmacia, UK), following the two-step labelling method of Schenk et al. (2000). The dye labelling of the control and test samples were swapped among the replicate experiments to minimise any bias in cDNA incorporation and photo-bleaching of the fluorescent dyes. The pre-hybridisation of the microarray slides, hybridisation of the target cDNA and subsequent washing of the slides to remove unbound target was performed following the protocol for the CMT-GAPS-coated microscope slides provided by the manufacturer (Corning, USA). The slides were scanned with a GenePix 4000 A microarray scanner (Axon Instruments, Union, Calif.). The features were analysed using the GENEPIX PRO 4 software, and unsatisfactorily segmented features were either manually adjusted or discarded to ensure the integrity of the data obtained.

Statistical analysis of microarray data

The analysis of the microarray data files was carried out using functions contained in the TRMA statistical pack-

age. These functions operate as part of a statistical software package called R (http://www.r-project.org/). A detailed description of TRMA (tools for R microarray analysis; Wilson et al. 2003) is available online (http://www.csiro.au/gena/trma). The data sets generated from the GENEPIX software were loaded into TRMA using the LOADGENEPIXFILE function. Normalisation of $\log_2$ ratio values was performed using the SPATIALLYNORMALISE function. This method of normalisation corrected for spatial and intensity-dependent effects of fluorescence across the microarray slide (Wilson et al. 2003). In addition, the possible biases in fluorescence due to differences in the efficiency of incorporation of the two dyes and unequal loading of cDNA samples were also corrected. Using the median values of the normalised log ratios for each gene in each replicate, we determined differentially expressed genes using the FINDDIFFEXPGENES function. Differentially expressed genes are selected as outliers in a Gaussian distribution of the entire data set. Therefore, a ratio cut-off was empirically computed from the normalised $\log_2$ ratio data, which were rescaled and centred in order to make direct comparisons between slides in all four replicates. The COMPAREINTERESTINGGENES function was used to compare the differentially expressed genes among the replicates. The stringency was set to 100%—i.e. only those genes differentially expressed in all of the replicates tested, with a consistent direction of expression were selected.

Microarray analysis of wheat cultivars according to the *Pin* mutation causing hardness

Endosperm from seed at 9–10 DPA was compared for the hard-seed cultivars, Falcon and Hartog, which have the null *Pina* mutation and for Cook, Janz and Sunstar, which have a mutation in the *Pinb* gene, to the following soft cultivars: Heron, Corrigin, Kulin and Egret. Using the NSWCAG array the following comparisons were made: (1) for the null *Pina* mutation, Heron versus Falcon (dye swap); Heron versus Falcon (no dye swap); Kulin versus Hartog (no dye swap); (2) for those lines carrying a mutation in the *Pinb* gene, Corrigin versus Cook (dye swap); Heron versus Cook (no dye swap); Kulin versus Janz (dye swap); Egret versus Sunstar (dye swap). For all of these comparisons the hard line is considered the control and labelled with Cy3; this is listed as no dye swap. If the soft wheat is labelled with Cy3, it is listed as a dye swap.

Sequence confirmation of clones

To confirm the identity of the differentially expressed genes, these genes were sequenced from the 5′ ends of the cDNA clones using the M13/pUC reverse primer and Big Dye Terminator Cycle Sequencing (ABI, Foster City, Calif.) with 0.12 μg of the PCR-amplified insert from the cDNA clone as the template.

Analysis of RNA

The RNA was extracted from the caryopsis and 10 μg of total RNA was separated on a 1.4% agarose-formaldehyde gel (w/v) and transferred to Hybond N+ membrane (Amersham) using standard alkaline transfer protocols (Sambrook et al. 1989). Hybridisation of probes was at 65°C in Khandjian hybridisation buffer (Khandjian 1987). The membrane was washed once for 30 min at 65°C with 2× SSC, 0.1% (w/v) sodium dodecyl sulphate (SDS) and 2× 40 min at 65°C with 0.2× SSC, 0.1% (w/v) SDS and then exposed overnight at 80°C with Kodak MR X-ray film using an intensifying screen.

Probes used for hybridisation

Probes used for the RNA hybridisations were made by amplifying the inserts from the cDNA clones. The PCR SuperMix (Invitrogen, Carlsbad, Calif.) was used with 3 ρmol of each primer and 50 ng template DNA. The probes were amplified under the following conditions: one cycle of 94°C for 5 min; 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 2 min; one cycle of 25°C for 1 min. The inserts were labelled using the MegaprimeDNA labelling system (Amersham Biosciences, Piscataway, N.J.).

Quantitative real-time PCR

The level of differential gene expression observed from the microarray was verified by quantitative real-time PCR (RT-PCR) using a Rotor-Gene 2000 (Corbett Research, Mortlake, Australia). An alpha-tubulin gene was used as a control. The alpha-tubulin gene was selected from the microarray data using the SELECTHOUSKEEPINGGENES function in TRMA. This programme selects genes with a low differential expression and a low variability across the slides.

For the analysis, 1 μg of total RNA was used to produce cDNA using the Access RT-PCR System (Promega, Madison, Wis.), and 25 ng was used as a

Table 1 Primers used for the RT-PCR analysis

Gene to amplify	Primers[a]	Primers (5′ → 3′)[a]
Alpa thionin	F	GCAAGATAGCCTGCCAGC
	R	CCTTAGGGCAGCTTAGGC
Gsp-1	F	GCGATCTAAGTGGCTTCAAG
	R	GCTAGTGATGGGGATGTTGC
Pinb	F	ATGAAGACCTTATTCCTCCTA
	R	TCACCAGTAATAGCCACTAGGGAA
Alpha-tubulin	F	GGCTTGTGTCTCAGGTTATCTCATC
	R	CATGGAGGATGGCTCGAAGG
Pina	F	ATGAAGGCCCTCTTCCTCATAGG
	R	TCACCAGTAATAGGCAATAGTGCC

[a]Primers: F, forward; R, reverse. The primers for *Pinb* are from Gautier et al. (1994)

template for RT-PCR following the method described in Jan Klok et al. (2002). Gene-specific primers were designed using sequence information from the arrayed genes to give a T_m between 55°C and 60°C and a product length of 150–300 bp. The primers used to amplify the selected genes are listed in Table 1.

Results

Comparison of Heron NIL

For the first analysis, four replicate microarray slides were compared at each of three time points–6, 15 and 25 DPA. The comparison was made between the hard and soft seeds, and the dyes used to label these phenotypes were swapped for two of the replicates at each time point. The microarray results were analysed using TRMA, and the genes identified as significantly differentially expressed in all four replicate samples were identified (Table 2). Two clones are listed as mixed in Table 2. Sequence analysis and hybridisation of these clones showed that the hybridising fragment on the array corresponded to hordoindoline-a (the barley equivalent of puroindoline-a) in both cases, indicating that these array features had been incorrectly annotated in the database.

At the early time point, there were no transcriptional changes that were consistent among all replicates. At the middle and latter stages, the only genes identified were *Pina* and *Pinb*, which were up-regulated in soft NILs. As there were no genes differentially expressed in all four replicates at the early stage, these slides were analysed under less stringent conditions. The comparisons were made between two dye-swapped replicates at 100% stringency. From the four possible pair-wise comparisons, a total of 48 differentially expressed genes were identified. Using the GENEPIX software, we examined the fluorescent intensity of all these features on the individual microarrays. Examination of the individual slides revealed that in two of the replicates there was increased expression in the hard seed for these genes and that in the other two the expression was decreased. Because of this inconsistent expression, ten genes were selected from the 48 genes identified and their level of expression in both hard and soft seeds determined using a Northern analysis. For the Northern analysis, RNA was isolated from fresh replicate samples of 6-DPA endosperm tissue. The Northern results showed that in the majority of cases these ten genes did not follow the expression pattern obtained from the two array comparisons (Table 3), thereby confirming the variable expression observed among the four microarray slides.

Table 2 TRMA results from the microarray analysis of the Heron NIL comparison using four replicates

Identification	Name	Log_2[a]	Standard deviation
Middle: 15 DPA			
BQ606719	Puroindoline-a (*Pina*)	4.16	0.35
CV065963	Puroindoline-b (*Pinb*)	1.16	0.09
MIXED[b]	Hordoindoline-a (*Pina*)	3.4	0.55
MIXED	Hordoindoline-a (*Pina*)	2.84	0.23
BQ605639	Puroindoline-b (*Pinb*)	0.92	0.19
Late: 25 DPA			
BQ606719	Puroindoline-a (*Pina*)	3.41	1.54
MIXED	Hordoindoline-a (*Pina*)	2.84	1.61
MIXED	Hordoindoline-a (*Pina*)	2.32	1.06

[a]The data are expressed as Log to base 2 transformed ratios; positive values indicate up-regulation of the gene in soft seed
[b]MIXED indicates that two different genes were at the same location on the array; the gene listed is the one that hybridised to the probe

Comparison of Falcon NIL

For this analysis, two replicates with a dye swap were examined, and the results are listed in Table 4. The clones listed as mixed in this table were shown by sequence analysis and hybridisation to be hordoindoline-a (see above).

As in the case for the Heron NIL, the main gene differentially expressed at both the middle and late stages was *Pina*, up-regulated in soft seed, but in this comparison there was no change for *Pinb*. One gene, a Bowman-Birk trypsin inhibitor, was down-regulated in the soft seed at the late developmental stage. At the early developmental stage, all of the differentially expressed genes, which included three alpha-purothionins and one grain softness protein, were down-regulated in the soft NILs. There were no genes in common between the genes seen in the TRMA comparisons using two replicates of the Heron NILs, mentioned above, and those seen here for the Falcon comparison.

Table 3 TRMA results comparing the microarray analysis of the Heron NIL, two replicates, using RNA from 6-DPA endosperm and the results from a Northern analysis of these genes

Identification	Gene	Nonscaled[a]	Northern analysis[b]
BQ605601	Protease inhibitor	2.7	Down, soft
AW448662	Cell division protein	2.1	Equal
BQ609053	Hypothetical protein AL022198	2	Up, soft
BQ607333	TRFA (AB009080)	1.8	Equal
BQ609037	Gama-gliadin	1.7	Equal, but extra band in soft
BQ605797	Glutaminyl tRNA synthase	2.5	Equal
BQ609138	α/β-gliadin	1.5	Equal
BQ607866	Sucrose synthase	2.1	Up, soft
CV066675	Putative protein (AL035602)	1.5	Down, soft
CV064591	Putative blue copper (AC003105)	1.7	No hybridisation

[a]The nonscaled ratio gives the actual difference in expression between soft and hard seeds as observed on the microarrays. Positive values indicate up-regulation of the gene in soft seed
[b]The description indicates the level of expression observed for the probe when comparing the hybridisation intensity from the RNA of the soft and hard seed

Table 4 TRMA results from the microarray analysis of the Falcon NIL comparing two replicate arrays

Identification	Name	Log_2[a]	Standard deviation
Early			
BQ606978	Alpha-purothionin	−1.08	0.00
BQ607864	Alpha 2 purothionin	−1.27	0.08
BQ608643	Alpha-purothionin	−1.06	0.24
BQ609391	Grain softness protein (*GSP-1*)	−1.25	0.17
Middle			
MIXED[b]	Hordoindoline-a (*Pina*)	2.40	0.17
MIXED	Hordoindoline-a (*Pina*)	2.63	0.13
BQ606719	Puroindoline-a (*Pina*)	5.87	0.10
Late			
MIXED	Hordoindoline-a (*Pina*)	1.35	0.04
MIXED	Hordoindoline-a (*Pina*)	1.25	0.21
BQ606719	Puroindline-a (*Pina*)	1.84	0.32
BQ608545	Bowman Birk trypsin inhibitor	−1.78	0.72

[a]The data are expressed as Log to base 2 transformed ratios; positive values indicate up-regulation of the gene in soft seed
[b]MIXED indicates that two different genes were at the same location on the array; the gene listed is the one that hybridised to the probe

Table 5 TRMA results from the microarray analysis of the *Pina* null cultivars

Identification	Name	Log_2[a]	Standard deviation
BF428929	Puroindoline-a (*Pina*)	2.75	0.86
BQ606719	Puroindoline-a (*Pina*)	1.97	0.74
BQ608212	HSP 70 (AF005993)	1.05	0.38
CV066040	No hit[c]	0.88	0.16
UNKNOWN[b] 8101 8104 9 23	UNKNOWN	−0.82	0.23
AL811585	proteasome Z subunit	−0.87	0.42
BQ605670	No hit	−1.04	0.52
BQ609159	Bril-associated receptor-like kinase	−1.37	0.46
BF424643	No hit	−3.04	1.23

[a]The data are expressed as Log to base 2 transformed ratios; positive values indicate up-regulation of the gene in soft seed
[b]UNKNOWN indicates that no sequence was obtained for this feature on the array
[c]No hit means that no match was found for the gene sequence of that feature with either a BLASTX or BLASTN search

A comparison of wheats containing the *Pinb* and null *Pina* mutation

These comparisons were made between soft and hard wheats, where the hardness of the seed was attributed to either the *Pina* null or the *Pinb* mutation. The material labelled for the array was total RNA obtained from 9- to 10-DPA endosperm tissue. The comparisons made were those listed in the Materials and methods section. As with the previous experiment, the hard line was considered to be the control and labelled with Cy3 (no dye swap). These comparisons were made using the NSW-CAG array, which contains the same endosperm clones as present on the array used for the Heron and Falcon NIL experiments but with an additional 11,000 clones from a range of tissues (see Materials and methods).

For the two *Pina* null lines, those genes differentially expressed in both comparisons are listed in Table 5. The major change was associated with expression of *Pina*, which was up-regulated in soft seeds. Several other genes are listed in Table 5 which had an approximate twofold up or down difference in expression level compared to the hard seed. Not one of these clones, however, was identified in the Heron and Falcon NIL comparisons. Following sequence verification of these clones and a blast comparison, the majority were unknown sequences. However, as these clones were identified from only two replicates, and given the variability of gene expression seen at this early development stage, we did not characterise them any further.

When lines with the *Pinb* mutations were compared with soft wheats (four microarrays at 100% stringency) no genes were identified as being differentially expressed at 9–10 DPA.

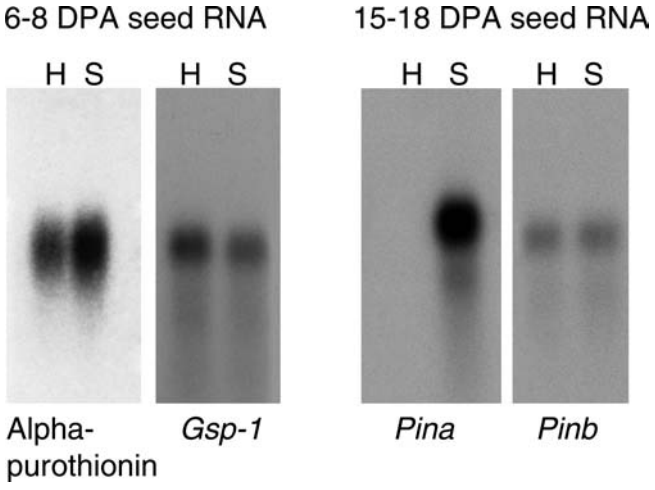

Fig. 1 Total RNA from wheat cv. Heron hybridised, *from left to right*, with alpha-purothionin, *Gsp-1*, *Pina* and *Pinb* probes. The time of exposure of the autoradiographs varied for each probe. The difference in the level of endosperm expression for these genes between the NILs of Heron is shown. *H* Hard seed, *S* soft seed RNA

The complete data sets used for these analyses are on the GENA database at http://www.pi.csiro.au/gena/. All of the clones identified on the array were sequence-verified from the cDNA used for the microarray.

Northern analysis to verify expression

Total RNA was prepared from fresh batches of seed from the soft and hard Heron NIL. For the Northern analysis, RNA from the same developmental stage as that used in the microarray experiments was used for validation of the genes: from the early stage, alpha-purothionin and *GSP-1*; from the middle stage, *Pina* and *Pinb*. At the early developmental stage, *GSP-1* was

Table 6 Quantitative RT-PCR results[a] for genes identified in the microarray experiments as being differentially expressed. The results represent average values from four repeated experiments

	Heron NIL		Falcon NIL	
	Early[b]	Middle[b]	Early	Middle
Alpha-thionin	1.3	1.2	1.1	1.7
Pinb	4	1.5	1.1	2.1
Gsp-1	3	0.6	1.0	2.1

[a]Ratio values are given: positive values indicate up-regulation of the gene in soft seed
[b]Early RNA is from 6-DPA endosperm tissue; Middle RNA is from 15-DPA endosperm tissue

up-regulated and alpha-purothionin was down-regulated in hard seed. At the middle stage, *Pina* was only expressed in the soft seed, and *Pinb* was equally expressed in both types of seed (Fig. 1). These results confirmed the expression pattern observed on the microarray, except for the alpha-purothionin gene which showed the opposite expression pattern.

RT-reverse PCR

The expression changes observed on the Northern analysis were quantified using RT-PCR for *Pina* and *Pinb*, *Gsp-1* and the alpha-purothionin genes. For this analysis, RNA was extracted from both the Heron and Falcon NILs at the early and middle developmental stages. The RT-PCR reaction was repeated four times for each gene using the same RNA sample. The results for the *Pina* gene are not presented because no product was amplified from the hard seed and it was consequently not possible to give a ratio value. Also, at the early stage using the Heron hard seed RNA, the *Pinb* and *Gsp-1* genes gave practically no measurable product, and this skewed the results towards soft seed expression (Table 6).

Discussion

Microarrays offer the opportunity to directly compare the gene transcription profile of a ''test'' and a ''control'' phenotype and identify those genes that are associated with the difference in phenotype. Microarray comparisons are limited, however, by the gene coverage of the species under investigation and the level of differential gene expression that can be accurately determined.

In the first study presented here we evaluated the differences in gene transcription between soft and hard seed NIL of wheat at three time points. The genes used for this analysis were approximately 5000 unique cDNA clones isolated from wheat and barley endosperm tissue. An evaluation by Clarke et al. (2000) has demonstrated that 4000–8000 genes would be expressed in developing endosperm at an early to middle grain filling stage.

The results for the Heron and Falcon NILs show that the *Pin* genes were not differentially expressed at the early stage of grain filling—6 DPA—in either the Falcon or Heron background. At 15 DPA, the *Pina* and *Pinb* genes showed increased expression in the soft grain of Heron, and this persisted into the late stage of development, 25 DPA. In the Falcon background, only *Pina* was differentially expressed at the middle and late developmental stages; the expression of *Pinb* at these same time points was inconsistent among the replicate microarrays tested. In the RT-PCR analysis of the Heron and Falcon samples, at mid-endosperm development, a 1.5–2-fold increase in soft seed expression was demonstrated for the *Pinb* gene. These findings are similar to those of Turnbull et al. (2002) who measured the amount of the puroindoline proteins. Using enzyme-linked immuno-absorbent assay (ELISA), they showed that at 10 DPA they were unable to detect PINA protein in either the soft or hard grains of the Falcon or Heron lines. From 15 DPA onwards, there was a 30-fold increase in the amount of PINA and about a twofold increase in the amount of PINB in the soft grain and this difference was maintained until grain maturity.

Although *Pinb* has been associated with grain softness in many studies (Beecher et al. 2002; Krishnamurthy and Giroux 2001), the transcript abundance did not vary significantly between soft and hard seeds in any of the wheats tested here. These observations and the absence of any other significant transcriptional changes would lead one to conclude that the *Pin* genes are the primary cause of grain softness and that both these genes need to be present for the grain texture to be soft. The model proposed by Giroux and Morris (1998) in which they speculate that both puroindoline proteins need to be present and functional in order for the grain to be soft is in agreement with this observation. A more recent study by Capparelli et al. (2003) also demonstrated that the level of friabilin associated with starch is regulated by the level of expression of the *Pina* transcript. These researchers concluded that for fully functional friabilin (which determines grain softness) both PINA and PINB are required. In contrast, for rice it has been shown that the expression of *Pinb* alone is sufficient to bring about increases in grain softness (Krishnamurthy and Giroux 2001). This observation represents an important difference between the model proposed for wheat and experimental results reported for rice. These differences could be due to the very different composition of the rice and wheat endosperm (Kavakli et al. 2000).

In this study, we observed inconsistent gene expression among the replicates at the early stage of endosperm development. At this early stage, differential gene expression was identified for a variety of genes in the TRMA analysis, but only when two biological replicates were compared. A Northern analysis of several of these clones on RNA isolated from new plant material confirmed the inconsistent expression patterns at this early stage of endosperm development (Table 3). This incon-

sistent gene expression profile observed among the replicates at 6 DPA probably means that the gene changes observed are not involved with grain texture but that they are more likely associated with differences in grain development. Although every effort was made to select only seeds for each replicate at the same developmental stage, inevitably 1-day to 2-day differences in development may have been missed. This variation in seed development is especially critical for the early time points when so many changes are taking place in a relatively short time frame. For instance, between 1 DPA and 5 DPA the seed progresses from fertilisation through nuclear division to form a multinucleate cytoplasmic matrix around the periphery of the central cell. Cell-wall formation then takes place to form large vacuolate cells (Simmonds and O'Brien 1981). During these rapid changes, many genes would be expected to be active for short periods of time. The fact that the genes observed as being differentially expressed in the early Falcon NIL comparison were not the same as those observed in the early Heron NIL comparisons tends to give weight to this conclusion.

In a second study, we evaluated differences in transcription between a range of soft and hard wheat cultivars which either had the *Pina* null or a *Pinb* mutation causing hardness. For this study, the array contained the same 5,000 endosperm clones as used in the first study but with an additional 11,000 unique sequences from a range of wheat tissue types. The logic behind this comparison was the assumption that the underlying mechanism influencing grain texture would be the same among the cultivars compared, the *Pina* null or the *Pinb* mutation. It was assumed that gene expression representing cultivar differences would be removed by comparing the arrays in the analysis at 100% stringency—i.e. only those genes that occurred in all of the arrays would be considered to be differentially expressed. With the *Pina* null comparison, only two cultivars were available and, hence, the large number of additional genes observed. However, with the exception of *Pina*, none of the other genes identified were in either the Heron or Falcon NIL comparisons, Therefore, it was assumed that these genes were not linked to grain hardness. No genes were identified as being differentially expressed in the *Pinb* null comparison.

Our results indicate that it is necessary to do more biological replication at early developmental time points in order to differentiate between those genes that are involved in the trait of interest and those that are involved in development. A greater level of replication, however, could mask those genes active for short time periods, unless the developmental stages of the biological replicates are accurately synchronised.

Another complicating factor for wheat microarray experiments is the hexaploid nature of its genome. With three homoeologous genomes and many multi-gene families, most genes on the array are the target for at least three and possibly more related genes from the test and control RNA. These genes, although related by sequence similarity, could carry out different but related functions in the plant. However, the sequence similarity could lead to cross-hybridisation, which would mask or distort the ratio of differential gene expression observed. The hybridisation conditions that we used in the microarray experiments allow genes with an 80% sequence similarity to anneal. Therefore, cross-hybridisation may occur among gene family members. This cross-hybridisation may partly explain the differences observed between the microarray and the Northern analysis because in the Northern analysis more stringent hybridisation conditions were used to discriminate between the target and probe sequences. Another reason for the differences could be that RNA from slightly different developmental time points was evaluated in the Northern analyses than in the microarray (as discussed above).

The similarity of the genomic DNA between the hard and soft seed in both the Heron and Falcon backgrounds is evident from the small number of genes listed as being differentially expressed in this microarray analysis. These results suggest that it is the presence of functional *Pina* and *Pinb* genes, rather than their expression level, which is the major determining factor in grain texture. In addition, if other genes are involved in the hardness character their function is not influenced by transcriptional regulation, or the genes were not present on the microarray.

References

Ali S, Holloway B, Taylor WC (2000) Normalisation of cereal endosperm EST libraries for structural and functional genomic analysis. Plant Mol Biol Rep 18:123–132

Bechtel DB, Wilson JD (1997) Ultrastructure of developing hard and soft red winter wheats after air- and freeze-drying and its relationship to endosperm texture. Cereal Chem 74:235–241

Bechtel DB, Wilson JD, Martin CR (1996) Determining endosperm texture of developing hard and soft red winter wheats dried by different methods using the single-kernel wheat characterisation system. Cereal Chem 73:567–570

Beecher B, Bettege A, Smidansky E, Giroux MJ (2002) Expression of wild-type *pinB* sequence in transgenic wheat complements a hard phenotype. Theor Appl Genet 105:870–877

Blochet JE, Chevalier C, Forest E, Pebay-Peyroula E, Gautier MF, Joudrier P, Pezolet M, Marion D (1993) Complete amino acid sequence of puroindoline, a new basic and cystine-rich protein with a unique tryptophane-rich domain, isolated from wheat endosperm by Triton X114 phase partitioning. FEBS Lett 329:336–340

Capparelli R, Boriello G, Giroux MJ, Amoroso MG (2003) Puroindoline A-gene expression is involved in association of puroindolines to starch. Theor Appl Genet 107:1463–1468

Chantret N, Cenci A, Sabot F, Anderson O, Dubcovsky J (2004) Sequencing of the *Triticum monococcum* Hardness locus reveals good microcolinearity with rice. Mol Gen Genomics 271:377–386

Clarke BC, Hobbs M, Skylas D, Appels R (2000) Gene active in developing wheat endosperm. Funct Integr Genomics 1:44–55

Gautier MF, Aleman ME, Guirao A, Marion D, Joudrier P (1994) *Triticum aestivum* puroindolines, two basic cysteine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol Biol 25:43–57

Giroux MJ, Morris CF (1997) A glycine to serine change in puroindoline b is associated with wheat grain hardness and low levels of starch-surface friabilin. Theor Appl Genet 95:857–864

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

Giroux MJ, Talbert L, Habernicht DK, Lanning S, Hempill A, Martin JM (2000) Association of puroindoline sequence type and grain hardness in hard red spring wheat. Crop Sci 40:370–374

Greenblatt GA, Bettge AD, Morris CF (1995) Relationship between endosperm texture and the occurrence of Friabilin and bound polar lipids on wheat starch. Cereal Chem 72:172–176

Greenwell P, Schofield JD (1986) A starch granule protein associated with endosperm softness in wheat. Cereal Chem 63:379–380

Greenwell P, Schofield JD (1989) The chemical basis of grain softness and hardness. In: Wheat end-use properties. Proc Int Assoc Cereal Sci Technol Symp. Helsinki Press, Helsinki, pp 59–71

Higgins TJ, Zwar JA, Jacobsen JN (1976) Gibberellic acid enhances the level of translatable mRNA for alpha-amylase in barley aleurone layers. Nature 260:166–168

Jan Klok E, Wilson IW, Wilson D, Chapman SC, Ewing RM, Sommerville SC, Peacock WJ, Dolferus R, Dennis E (2002) Expression profile analysis of the low-oxygen response in Arabidopsis root cultures. Plant Cell 14:2481–2494

Jolly CJ, Rahman S, Kortt AA, Higgins TJV (1993) Characterisation of the wheat Mr 15,000 'grain softness protein' and analysis of the relationship between it's accumulation in the whole seed and grain softness. Theor Appl Genet 86:589–597

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

Kavakli IH, Slattery JC, Ito H, Okita TW (2000) The conversion of carbon and nitrogen into starch and protein in developing storage organs: an overview. Aust J Plant Physiol 27:561–570

Khandjian EW (1987) Optimised hybridisation of DNA blotted and fixed to nitrocellulose and nylon membranes. Biotechnology 5:165–167

Krishnamurthy K, Giroux MJ (2001) Expression of wheat puroindoline genes in transgenic rice enhances grain softness. Nat Biotechnol 19:162–166

Mattern PJ, Morris R, Schmidt JW, Johnson VA (1973) Locations of genes for kernel properties in the wheat variety Cheyene using chromosome substitution lines. In: Sears ER, Sears LMS (eds) Proc 4th Int Wheat Genet Symp. University of Missouri, Columbia, pp 703–707

Morris CF (2002) Puroindolines: the molecular genetic basis of wheat grain hardness. Plant Mol Biol 48:633–647

Oda S, Schofield JD (1997) Characterisation of friabilin polypeptides. J Cereal Sci 26:29–36

Perretant MR, Cadalen T, Charmet G, Sourdille P, Nicolas P, Boeuf C, Tixier MH, Branlard G, Bernard S, Bernard M (2000) QTL analysis of bread-making quality in wheat using a doubled haploid population. Theor Appl Genet 100:1167–1175

Rahman S, Jolly CJ, Skerritt JH, Wallosheck A (1994) Cloning of a wheat 15-kDa grain softness protein (GSP) GSP is a mixture of puroindoline-like polypeptides. Eur J Biochem 223:917–925

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM (2000) Coordinated plant defence responses in Arabidopsis revealed by microarray analysis. Proc Natl Acad Sci USA 97:11655–11660

Simmonds DH, O'Brien TP (1981) Morphological and biochemical development of the wheat endosperm. In: Pomeranz Y (ed) Advances in cereal science and technology, vol 4. American Association of Cereal Chemists, St. Paul, Minn., pp 5–70

Sourdille P, Perretant MR, Charmet G, Leroy P, Gautier MF, Jourdier P, Nelson JC, Sorrels ME, Bernard M (1996) Linkage between RFLP markers and genes affecting kernel hardness in wheat. Theor Appl Genet 93:580–586

Staden R, Beal KF, Bonfield JK (1998) The Staden package. Computer methods in molecular biology. In: Misener S, Krawetz SA (eds) Bioinformatics methods and protocols, vol 132. Humana Pres, Totowa, pp 115–130

Symes KJ (1961) Classification of Australian wheat varieties based on the granularity of their wholemeal. Aust J Exp Agric Anim Husbandry 1:18–23

Symes KJ (1965) The inheritance of grain hardness in wheat as measured by the particle size index. Aust J Agric Res 16:113–123

Symes KJ (1969) Influence of a gene causing hardness on the milling and baking quality of two wheats. Aust J Agric Res 20:971–979

Tranquilli G, Lijavetzky D, Muzzi G, Dubcovsky J (1999) Genetic and physical characterisation of grain texture-related loci in diploid wheat. Mol Gen Genet 262:846–850

Turnbull KM, Rahman S (2002) Endosperm texture in wheat. J Cereal Sci 36:327–337

Turnbull K-M, Marion D, Gaborit T, Appels R, Rahman S (2002) Temporal expression of grain hardness in the developing wheat endosperm. Planta 216:699–706

Turnbull K-M, Turner M, Mukai Y, Yamamoto M, Morell MK, Appels R, Rahman S (2003) The organisation of genes tightly linked to the *Ha* locus in *Aegilops Tauschii*, the D genome donor to wheat. Genome 46:330–336

Wilson DL, Buckley MJ, Helliwell CA, Wilson IW (2003) New normalisation methods for cDNA microarray data. Bioinformatics 19:1325–1332