

A microarray-based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat

Per L. Gregersen*, Henrik Brinch-Pedersen & Preben B. Holm

Department of Genetics and Biotechnology, Danish Institute of Agricultural Sciences, Research Centre Flakkebjerg, Denmark, Slagelse

Received 8 September 2004; accepted 1 August 2005

Key words: cDNA microarrays, gene expression, phytase, safety assessment, transgenic wheat

Abstract

Global, comparative gene expression analysis is potentially a very powerful tool in the safety assessment of transgenic plants since it allows for the detection of differences in gene expression patterns between a transgenic line and the mother variety. In the present study, we compared the gene expression profile in developing seeds of wild type wheat and wheat transformed for endosperm-specific expression of an *Aspergillus fumigatus* phytase. High-level expression of the phytase gene was ensured by codon modification towards the prevalent codon usage of wheat genes and by using the wheat 1DX5HMW glutenin promoter for driving transgene expression. A 9K wheat unigene cDNA microarray was produced from cDNA libraries prepared mainly from developing wheat seed. The arrays were hybridised to fluorescently labelled cDNA prepared from developing seeds of the transgenic wheat line and the mother variety, Bobwhite, at three developmental stages. Comparisons and statistical analyses of the gene expression profiles of the transgenic line vs. that of the mother line revealed only slight differences at the three developmental stages. In the few cases where differential expression was indicated by the statistical analysis it was primarily genes that were strongly expressed over a shorter interval of seed development such as genes encoding storage proteins. Accordingly, we interpret these differences in gene expression levels to result from minor asynchrony in seed development between the transgenic line and the mother line. In support of this, real time PCR validation of results from selected genes at the late developmental stage could not confirm differential expression of these genes. We conclude that the expression of the codon-modified *A. fumigatus* phytase gene in the wheat seed had no significant effects on the overall gene expression patterns in the developing seed.

Introduction

Integration of transgenes into plant genomes may potentially lead to inactivation or alteration of endogenous gene activities. An often expressed concern is that transgene integration or transgene expression may cause changes in cellular metabolism that are not detected during the agronomic

evaluation and conventional safety assessment routines for transgenic plants to be used as feed or food. In theory, such changes in metabolism could constitute a potential risk to the health of livestock and man.

The development of global profiling techniques provides opportunities to address the questions raised by the concerns about unintended side effects in transgenic plants. Via profiling of plant compounds these techniques allow for global assessments of potential changes in the metabolism

*Author for correspondence
E-mail: per.gregersen@agrsci.dk

of transformed plants (Kuiper et al., 2003). These techniques could be implemented as strong experimental and safety assessment tools during the development and testing of transgenic plants. Profiling techniques thus carry the promise of being robust and unbiased methods to compare the transgenic plants to their non-transgenic counterparts in order to reveal any unintended effects of the transgene insertion (Kok & Kuiper, 2003; Cellini et al., 2004). This comparative approach, together with specific knowledge about the inserted transgene, its encoded product, and the site of integration can form the basis for further analysis on aspects related to toxicology, allergy, and nutrition.

Several techniques are available for gene transcript, protein, and metabolite profiling (for a review, see Fiehn et al., 2001). Gene expression profiling is currently the most global tool since DNA microarrays comprising many thousand genes can be made for a range of species. However, while DNA microarrays are very powerful tools for identifying changes in gene expression they are only indirectly informative with regard to the cellular complement of primary and secondary metabolites. Here, proteome and metabolome analyses are required. These technologies still have a number of limitations in the form of resolution, identification of compounds, and biases in e.g. extractabilities (Fiehn et al., 2001; Trethewey, 2004). It is apparent though that gene expression profiling combined with focused studies of parts of the protein and metabolite complement is very informative and could be useful instruments in a safety assessment of transgenic plants (for discussion, see Cellini et al., 2004). In addition, this technology is constantly improved, e.g. with full coverage of transcriptomes of crop plants and with sophisticated statistical tools adapted to the analysis of the very complex data arising from the profiling techniques.

In this study we have compared the gene expression profiles of a transgenic wheat line that exhibits high expression level of a fungal phytase with that of the mother variety. We have focused on the seed since this is the plant product used for feed and food. The aim of using transgenic phytase cereals is to allow for an improved hydrolysis of the major phosphate reserve in seeds, phytic acid, in the digestive tract of monogastric animals. (Brinch-Pedersen et al.,

2000, 2002). In the present experiments we have monitored and compared the gene expression profiles in wild type and transgenic wheat at three different developmental stages of seed development. This allowed us to evaluate potential effects of the transgene on the overall gene expression profiles against the background of progressive developmental changes of the biological system in which the transgene is introduced.

The phytase construct was designed for endosperm-specific expression and targeting to the apoplast. This implies a localisation separate from that of its substrate, phytate, which is localised in the storage vacuoles of the aleurone layer (Simmonds & O'Brien, 1981). We have not performed localisation studies, but metal dye detection HPLC analysis of the myoinositol phosphate profile in wheat transformed with an *Aspergillus niger* phytase gene using the same promoter and signal sequence as in the present study did not reveal any differences compared to wild type wheat (Brinch-Pedersen et al., 2003). Thus, in theory the absence of phytate breakdown during grain development implies that any potential unintended effects of the phytase transgene should relate to the integration of the transgene by e.g. inactivation of endogenous genes or reactivation of dormant genes. We are aware, however, that other factors could play a role as well in this context such as a separate integration of the *bar* selectable marker gene, activity of the phosphinothricin acetyl transferase enzyme, and somaclonal variation generated during the tissue culture regimes. In addition, we cannot exclude that the heterologous phytase may exert some unspecific phosphatase activity and in that way affect cellular metabolism and gene expression.

One of the important aspects, when handling data from gene expression profiling experiments, is the use of statistical methods that allow for a comprehensive analysis of the data instead of using simple fold decreases or increases in gene expression levels in comparisons of mRNA samples (for reviews on microarray analysis, see e.g. Quackenbush, 2002; Smyth et al., 2003). We have in this study implemented tools from the Bioconductor package (www.bioconductor.org), which is a freely available package with a multitude of up-to-date tools for analysis of DNA microarray data.

Materials and methods

Cloning and codon optimisation of the Aspergillus fumigatus phyA phytase gene

The *Aspergillus fumigatus* strain ATCC 34625, obtained from the American Type Culture Collection, was propagated o/n at 30°C by shaking (250 rpm) in PYG 663 medium holding 1.25 g peptone/l, 1.25 g yeast extract/l, 3.0 g glucose/l and 20.0 g agar/l. Fungal mycelium was obtained after filtration through a 0.45-µm filter on a Erlenmeyer flask with suction (Millipore). Genomic DNA was prepared as previously described (Pasamontes et al., 1997). The *A. fumigatus phyA* phytase gene was PCR amplified as three fragments of 137, 532 and 750 bps, respectively. Together the three fragments covered the entire gene sequence. The reaction used BIO-X-ACT™ proof-reading DNA polymerase (Bioline, UK) and 200 ng genomic DNA. Codon optimisation and generation of restriction sites for further cloning were performed during the amplification. The fragments were cloned in pCR®II-TOPO vector (Invitrogen) and subsequently sequenced on an ABI Prism 310 Genetic Analyzer (Applied BioSystems). The primers introduced several changes to the native sequence, thus underlined nucleotides indicate restriction sites used for further cloning. Bold indicates that the nucleotide has been changed compared to the native sequence, and italics represent overhang for efficient digestion at the terminal restriction sites. The 137-bp fragment was amplified using the upper strand primer (1 *GGAGCGTCGACCTGCAGTCCAAGTCCTGC GACACGGTGGACCTCGGCTACCAGTGCT CCCCTGCGACCTCTCACCTCTGG* corresponding to bases 419–482 of the native *A. fumigatus* phytase gene and the lower strand primer #1 5'TGGGTACCGCGCGCCGTGGCGCGAGAG CACCTGCACCAAGGTGATCCGGCAGTCCT TGGGAAGCTTGCTCG spanning bases 528–599 of the phytase gene. The 532-bp fragment used the upper strand primer #2 5'GCGGTACCCA-ACCAGCTCCAAGAGCAAGAAGTACAAGA AGCTCGTGACGGCGATCCAGGCCAACGC CACCGACTTCAAGGGCAAGTTCGCCTTCT TG spanning bases 589–685 and the lower strand primer #2 5'CGCGAGCTCGGATGTCGGGCG

CGAAG AGCGCGGTGAAGTTGGCCGCCA CCTCGTCGCCAGCTGGCTCGCCTCGAAC TTGGTGC spanning bases 978–1062. The 750-bp fragment was amplified using the upper strand primer #3 5'CCGAGCTCGCGCCGAGAAGCA CCTCCCTGGCGTGACGCTGACCGACGAG GACGTGGTCAGCCTCATGGAC spanning bases 1051–1120 and the lower strand primer #3 5'G CTGGAATTGCGCCTTCGCGGATCCG CTGAAGCACTCGCCCCAGTTGCCCCCGGA CCTGGCCCAGCTCAAGCCCTTGACGAAG spanning bases 1678–1735 of the *A. fumigatus* gene. Following digestion, the three fragments (*SalI-KpnI*; *KpnI-SacI* and *SacI-EcoRI*) were ligated into plasmid pUC18-Δ*PstI* (pUC18 where the *PstI* site has been removed). The resulting plasmid pPhyFum-mod contained a full-length *A. fumigatus* phytase gene optimised for expression in wheat (GenBank accession # AY582135). The codon changes were designed to approximate the codon usage of highly expressed wheat genes (high percentage of XXG/C codons) and complied with the preferences suggested by the codon usage database (<http://www.kazusa.or.jp/codon/>). Consequently, the G/C content in the wobble base position was increased from 69.9% in the native to 79.5% in the modified sequence.

Construction of a plasmid for expression of codon optimised A. fumigatus phytase in wheat endosperm

A 80-bp barley α-amylase signal peptide was isolated as a *PstI-PstI* fragment from plasmid pUSPPhyN (Brinch-Pedersen et al., 2000) and ligated into the *PstI* site of pPhyFum-mod yielding pSPPhyFum-mod. A *SphI* and *SalI* fragment retaining 1251 bp of the wheat HMW (high molecular weight) glutenin 1DX5 promoter was isolated from plasmid pJD330 (kindly provided by Dr. J. Napier, Long Ashton Research Station, UK) and ligated into the *SphI* and *SalI* sites of pSPPhyFum-mod yielding p1DX5SPPhyFum-mod. Finally a 200-bp *BamHI-EcoRI* fragment holding the *nos* terminator sequence of the *Agrobacterium tumefaciens* nopaline synthase gene (Brinch-Pedersen et al., 2000) was ligated into p1DX5SPPhyFum-mod leading to p1DX5SPPhyFum-modN.

Generation of transgenic plants and phytase activity assays

Transgenic plants of the cultivar Bobwhite were generated by particle bombardment of immature embryos followed by selection on Bialaphos as previously described (Brinch-Pedersen et al., 1996, 2000). Embryos were bombarded with a 2:1 mixture of pUBARN- Δ dIII-HI:p1DX5SPPhy-Fum-mod-N. The *bar* selection gene (Thompson et al., 1987) in pUBARN- Δ dIII-HI was driven by the maize constitutive ubiquitin 1 promoter (Christensen et al., 1992). The presence of transgene inserts was in the first phase documented by PCR amplification using the upper strand primer 5'-GATGGCGAACAAACATTTGTCCCTCTCCCTC-3' corresponding to bases 1–31 of the α -amylase signal peptide (Rogers & Milliman, 1983) and the lower strand primer #1 already described. Subsequently, Southern blots were hybridised with a [α -³²]dCTP-labelled 333 bp *Sall*-*Bam*HI fragment of the PhyFum-mod gene. The procedures for the Southern blotting analysis as well as the phytase activity assays on flour prepared from mature grains are given in Brinch-Pedersen et al. (2000).

Fabrication of wheat cDNA microarrays

A 9K wheat unigene set was kindly made available by Keith Edwards, Bristol University (Wilson et al., 2004, www.cerealsdb.uk.net). This set of clones originates in 35 cDNA libraries prepared from a range of different tissues of the wheat plant but predominantly from tissues of the developing grain. Plasmid DNA mini-preparations were made from all clones using Millipore 96-well plates (MANNLY50, MAFBNOB50, Millipore) according to the manufacturer's recommendations. The cDNA inserts were PCR amplified from 0.5 to 1 μ l DNA mini-preparation using the T3 and T7 flanking primers (30 pmol) of the pBluescript cloning vector and *Taq* polymerase (0.8 U) (EP0402, Fermentas, LT) in 100 μ l reaction buffer (10 mM Tris, pH 8.8; 50 mM KCl; 0.08% Nonidet P40; 0.5 mM MgCl₂; 0.3 mM (each) dNTP). The PCR reactions were performed in a MJResearch Cycler (MJResearch) and the cycling conditions were 5 min at 96°C, followed by 45 cycles of 30 s at 96°C, 2 min at 58°C, 2 min at 72°C, and a final extension of 10 min at 72°C. PCR products were

precipitated with 2.5 vol 96% ethanol and 1/10 vol 3 M NaAcetate, pH 6.0, left at least 2 h at –20°C, and centrifuged 1 h, 4°C, 2600 rcf. The air-dried pellet was dissolved in 20 μ l of water. One μ l of each sample was run in a 1% agarose gel containing SYBR Green. The quality of the PCR products was visually inspected and abnormal or missing products were recorded. Five μ l (100–500 μ g) of each sample was vacuum-dried and re-dissolved in 5 μ l of 1 \times QMT spotting buffer (S201100, Quantifoil, D). The samples were spotted onto QMT amino slides (S122025, Quantifoil, D) using a Chipwriter microarrayspotter (Eurogentec, B) with 24 Stealth MP3 pins (Telechem Int., US). Following spotting, the slides were UV cross-linked at 250 mJ (Stratalinker, Stratagene) and kept dry and in the dark at room temperature until hybridisation.

Plant material for cDNA microarray hybridisations

Wheat plants of cv. Bobwhite and the transgenic line L07 were grown in soil plots in the greenhouse. Artificial illumination was used for supplementation and for ensuring a day/night cycle of 16/8 h. Individual spikes were tagged at pollination and developing grains harvested between 12 am and 2 pm at regular daily intervals following pollination. The grains were frozen in liquid nitrogen immediately after sampling and stored at –80°C until RNA extraction was performed.

RNA isolation and labelling of cDNA target preparations

For each RNA isolation, plant material from at least two different plants was combined, with 5–10 seeds per plant. Plant material was ground to a fine powder in liquid nitrogen using mortar and pestle, and RNA was isolated from approx. 2 \times 200 μ g of this material using oligodT₂₅-Dynabeads (610-05, Dynal, N), essentially following the recommendations of the manufacturer: First, the material was homogenised in 1.5 ml binding/lysis buffer (100 mM Tris-HCl pH 7.5, 0.5 M LiCl, 10 mM EDTA, 1% LiDS, 5 mM DDT) in 2 ml microtubes. Particulate material was sedimented by centrifugation (3–5 min) and the supernatant was mixed with beads from 200 μ l Dynabeads stock solution. Annealing of polyA⁺-RNA to the beads was allowed for 5–10 min at room temperature

followed by a wash in buffer A (10 mM Tris-HCl pH 7.5, 0.15 M LiCl, 1 mM EDTA, 0.1% LiDS). To increase the amount of annealed polyA⁺-RNA, the washed beads were mixed with a second batch of extraction from 200 µg ground plant material. Following the second annealing the beads were washed once in buffer A, twice in buffer B (10 mM Tris-HCl pH 7.5, 0.15 M LiCl, 1 mM EDTA), and three times in ice-cold RT wash buffer (55 mM Tris-HCl pH 8.3, 82.5 mM KCl). The beads were resuspended in 30 µl of reverse transcription reaction mixture (200 U SuperscriptII (InVitrogen), 40 U RNaseOUT (In-Vitrogen), 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl₂, 0.625 mM dNTP, 10 mM DDT). Following incubation for 1 h at 42°C, with the tubes placed horizontally in a rotary oven, the beads were washed twice in 300 µl TE (10:1) at room temperature and incubated 6 min at 95°C in 300 µl TE (10:1) to remove the RNA from the immobilised cDNA. For second strand synthesis, the cDNA beads were re-suspended in 24.5 µl of water and 4 µl of random nonamer primer solution (50 µM), incubated 3 min at 95°C, and annealed at room temperature 5–10 min. Then 12.6 µl second strand reaction mixture (10 U Klenow DNA polymerase I large fragment (EP0052, Fermentas, LT); 50 mM Tris-HCl, pH 8.0; 0.5 mM dATP, dCTP, and dGTP; 0.16 mM dTTP; 0.33 mM amino-allyl dUTP (A-21664, Molecular Probes, USA)) was added, and incubation was performed for 1 h at 37°C, with the tubes placed horizontally in a rotary oven to keep beads in suspension. The reaction mixture was discarded and the beads were washed twice in 300 µl wash buffer BX (0.1 M phosphate buffer pH 7.5, 1× SSC). Then the beads were resuspended in 40 µl of water and incubated 5 min at 95°C to elute second strand products. The supernatant was immediately recovered and kept on ice. The second strand synthesis from the immobilised cDNA on the beads was repeated and the cDNA concentration in the combined 80 µl amino-allyl-cDNA was estimated spectrophotometrically at 260 nm. Subsequently the product was vacuum dried (at 30°C) and re-dissolved in 10 µl of 0.1 M sodium bicarbonate buffer pH 9.0. The content of one vial of Cy3 or Cy5 NHS esters (RPN5661, Amersham Bioscience, UK) was dissolved in 2 µl of DMSO (D8418, Sigma), and the amino-allyl-cDNA was immediately added and mixed with the dye. The

mixture was incubated 1 h in the dark at room temperature, and subsequently 38 µl of 100 mM sodium acetate, pH 5.2, was added. From this solution the labelled cDNA was purified using the QiaQuick PCR purification kit (28104, Qiagen, D) according to the manufacturer's recommendations. Final elution from the columns was performed twice with 40 µl elution buffer. The full sample was used for spectrophotometrical estimation in a 50 µl quartz cuvette of the cDNA amount (from 260 nm readings) and Cy3 and Cy5 amounts (from 550 nm and 650 nm reading, respectively).

Hybridisation to spotted slides

Prior to hybridisation the spotted slides were blocked in succinic anhydride according to Eisen and Brown (1999). Following denaturation for 3 min in boiling water, the slides were then rinsed briefly in water and spin-dried in 50 ml tubes in a swing-out centrifuge at 1500 rpm for 6–8 min. The cDNA target solution was prepared from 25 pmol of Cy3- and Cy5-labelled cDNA, respectively, which were combined, vacuum dried, re-dissolved in 5 µl of water, and combined with 40 µl of QMT hybridisation solution (S230010, Quantifoil, D). Prior to addition to the slide, the target hybridisation solution was denatured for 2 min at 95°C. A LifterSlip (25x60I-2-4789, Erie Scientific Company, USA) was placed over the spotted area of the slide, and the target hybridisation solution was allowed to wick across the space created by the LifterSlip. The slide was placed in a closed hybridisation box containing a few ml of 0.5× SSC to keep high humidity, and incubation was performed for 16 h in a water bath at 42°C. The LifterSlip was removed in 2× SSC, 0.1% SDS, and the slides were washed once in 2× SSC, 0.2% SDS, preheated to 55°C, for 10 min; twice in 0.1× SSC, 0.1% SDS, preheated to 55°C, for 10 min; and twice in 0.1× SSC for 1–2 min. The slide was transferred to a 50 ml tube and immediately spin-dried in a swing-out centrifuge at 1500 rpm for 6–8 min.

Scanning and quantification of signals

Scanning was performed on an arrayWoRx microarray scanner (BioChipReader, Applied Precision, USA), which is a CCD camera-based microarray scanner. Exposure settings were optimised for each

slide. Quantification of hybridisation signals was performed using the arrayWoRx 2.0 Software Suite of the scanner. The spot grids for quantification were aligned manually with the spots for each slide. The raw data files were submitted to the ArrayExpress microarray data repository (<http://www.ebi.ac.uk/arrayexpress/>) along with details about experimental design and sample annotations.

Statistical analysis of microarray data

For diagnostic plots and statistical analysis of the cDNA microarray data the LIMMA software package (Smyth, 2003; Smyth & Speed 2003) from Bioconductor (www.bioconductor.org) was used. Appropriate filtering and normalisation preprocessing of the data were performed prior to the final analyses according to the help files of LIMMA and the LIMMA User's Guide included with the software. For the preprocessing of data for presentations of expression profiles over time normalisations of single channel signals were performed according to the LIMMA User's Guide.

Real time PCR validation

In order to validate microarray results for selected genes, real time PCR was performed using the ABIPrism7700 Sequence Detection System (Applied BioSystems). Optimal designs of primers for selected clones were made using the Primer Express software of the system. Total RNA was isolated from ground plant material using the Tri Reagent procedure (T9424, Sigma) according to the manufacturer's recommendations. For detection of PCR products in the real time PCR, the SYBR green master mix (4309155, Applied BioSystems) was used according to the manufacturer's recommendations. A dilution series of one of the samples was used to create standard curves for each of the primer sets that were subsequently used to determine the relative amounts of accumulating PCR products. Following normalisation to the 18S transcript, used as internal calibrator, the amounts of PCR products were used to calculate relative differences in expression levels among the samples.

Results

Transgenic wheat plants for microarray analysis

Transgenic plants from six individual lines were selected for further characterisation. The plants were fully fertile and looked phenotypically normal. Phytase activity measurements on T₁ seeds revealed significant increased phytase activities in all lines except line L05 (Figure 1). The highest activity was found in line L07 with 4571 FTU/kg compared to 742 FTU/kg in non-transgenic control seeds. Line L07 was selected for the cDNA microarray studies, and seeds harvested from T₂ plants were used for mRNA isolation in these studies. Southern blot analysis of line L07 (Figure 2) revealed the presence of only one integration site.

cDNA microarray fabrication

PCR products from a 9K wheat unigene set, acquired as bacterial stocks from Keith Edwards, Bristol Univ., were successfully produced from plasmid minipreparations. Quality checks in agarose gels revealed that 90% of the amplifications gave distinct bands. Gel photographs were filed as reference material for quality check of the final hybridisation results. The PCR products were spotted onto QMT amino slides. In total, each slide contained 48 pin groups with 19,968 elements, comprising duplicate spots of PCR products from 8959 proper genes and 1025 control or empty spots. The control spots comprised several replicates of different well-characterised plant genes and a number of Alexa488-labelled probes, used as navigation markers on the slides. Both types of control spots were used to estimate the quality of the spotting and hybridisations but were not included in the final statistical analysis of the results (see below).

Experimental design of hybridisation experiments

We designed an experimental setup that was balanced with respect to the use of Cy3 or Cy5 in the labelling reactions and which took into account the limitations in number of slides available (Figure 3). The strategy was to analyse differential gene expression between the two lines at different stages of the developmental process in

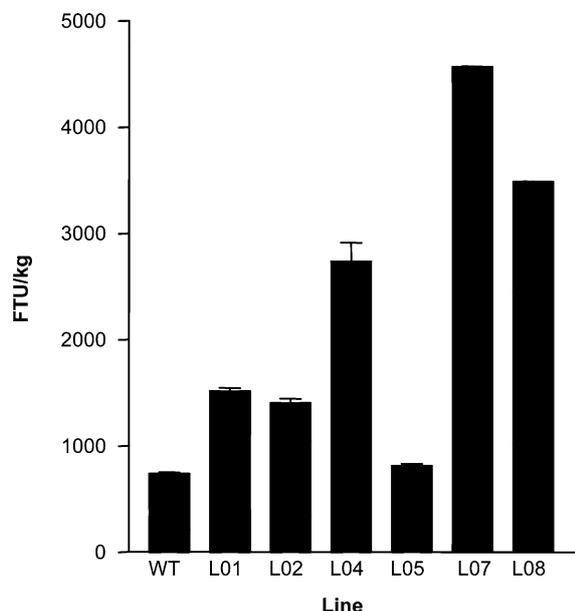


Figure 1. Phytase activity in mature T1 seeds of wild type and transgenic lines L01, L02, 04, L05, L07 and L08. One FTU is the amount of enzyme that liberates 1 μ mole orthophosphate per minute under the test conditions.

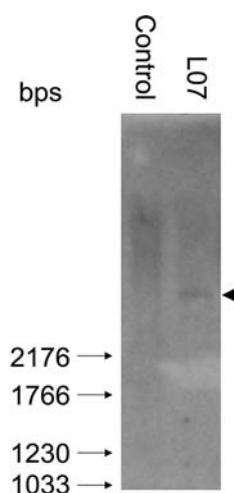


Figure 2. Southern blot analysis of *Bam*HI (single cut within plasmid) digested leaf genomic DNA (5 μ g) from control and line L07. Blots were hybridised with a 332 bp *Sall*-*Bam*HI fragment of the *PhyFum-mod* gene. An arrowhead indicates the position of a single hybridisation band in the L07 lane.

the seed and meanwhile to monitor the expression of the transgene. This was made possible by the inclusion of the transgene in the microarrays. In total, eight microarray hybridisations were performed in this design, with each sample type participating in at least two different hybridisa-

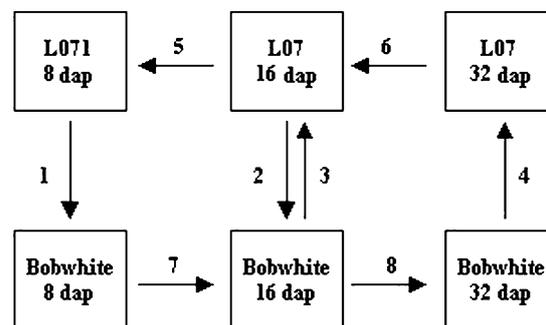


Figure 3. Experimental design of the cDNA microarray hybridisation experiments comparing gene expression levels in the transgenic line L07 and the control line Bobwhite. Boxes indicate RNA samples/plant sample origins. Independent samples were prepared for each hybridisation so that in total 16 samples were prepared, from harvest of plant material to cDNA labelling. Arrows indicate slides/microarrays with heads towards the samples labelled with Cy5 and tails at the samples labelled with Cy3. Abbrev: dap – days after pollination.

tions with balanced dye labelling. Plant material of the transgenic L07 line and the mother line was harvested at three different time points: early (7–9 days after pollination – designated 8 days after pollination (dap)), intermediate (15–18 days after pollination – designated 16 dap), and late 30–34 days after pollination – designated 32 dap). Individual hybridisation samples were prepared separately, i.e. from harvesting to hybridisation. Thus, at 8 and 32 dap two, and at 16 dap four biological replicates were used per sample origin. Since line L07 was still segregating for the transgene, material was only harvested from plants that were tested positive for the presence of the transgene using PCR. The phenotypic habitus of the two lines was indistinguishable. However, the development of the plants was not completely synchronous under the green house conditions, both within and between the two lines, and plant samples for the same developmental stage were accordingly in some cases harvested at different dates due to different timings in pollination. For the processing of samples for RNA isolation and labelling, ground grain material from at least two different plants were combined.

Microarray hybridisation and quantification of signals

The RNA isolation and labelling of cDNA relied on the use of oligodT₂₅-Dynabeads where 2nd

strand amino-allyl-cDNA was produced from immobilised 1st strand cDNA, using a random nonamer primer and Klenow DNA polymerase I fragment, and with subsequent coupling of Cy3 or Cy5 NHS esters to the cDNA. This indirect labelling procedure allowed for a monitoring of the labelling process through estimation of the amounts of produced cDNA and incorporated Cy-dyes, and it allowed for repeated reactions of the 2nd strand synthesis to increase the amounts of cDNA. Usually two rounds of 2nd strand synthesis of amino allyl-cDNA yielded sufficient amounts for one hybridisation, i.e. more than 25 pmol of incorporated Cy3 or Cy5 dye. The different RNA samples were labelled and hybridised pair-wise to the spotted microarray slides as indicated in the experimental design of Figure 3. Following hybridisation and washes the slides were scanned at the appropriate wave lengths for Cy3 and Cy5, and the signal intensities from individual spots were quantified using the arrayWoRx 2.0 Software Suite. Since the spot distribution was somewhat irregular across the slides and among slides, the spot grid for quantification was aligned manually to the spots for each individual slide. The extracted raw data for foreground and local background signals from the 8 hybridisations formed the basis for import into the R-based software tool LIMMA for further analysis. The raw data files are available from the ArrayExpress microarray data repository at EBI (<http://www.ebi.ac.uk/arrayexpress/accession: E-MEXP-368>) along with annotations about the hybridisation experiments.

Analysis of microarray data

In LIMMA the raw data was first subjected to diagnostic tools to evaluate the quality of the hybridisations. Diagnostic MA-plots (Dudoit et al., 2002) for whole arrays and for individual print-tip groups were produced and visually inspected for divergent distribution of spots. As an example, MA-plots for array 5, prior to and following print-tip-loess normalisation with background subtraction and omission of control spots, are shown in Figure 4. This demonstrates a reasonable distribution of spot intensities from background level to saturation in our arrays. The subtraction of background signals increases the spot variation, especially for weaker spots, and therefore the foreground signal alone was used for

normalisation and analysis of signals (see Alba et al., 2004). This probably causes an underestimation of the M-values from the analysis, but on the other hand it makes the data more robust for the analysis. In this context a special concern relates to spots holding the transgenes: since in theory no hybridisation signal should be obtained with non-transgenic control samples, it should not be possible to estimate an M-value for the transgenic vs. non-transgenic contrast. However, in practise there is a weak non-specific hybridisation signal from almost any spot (see Figure 5 for an example). By using this signal an M-value can be calculated that is, however, considerably underestimated. In this study, M-values of the transgenic vs. the mother line were calculated in this way and included in order to provide information about the relative expression levels of the transgenes in different samples.

Normalisation within the individual arrays was performed using the default print-tip-loess normalisation procedure of LIMMA. Box-plots of the signal distributions (M-values) (Yang et al., 2002) were produced to estimate differences in signal distributions across slides (data not shown). This indicated some variations as evidenced by higher spreads in intensity log ratios for some of the slides. This prompted implementation of the default scaling procedure for normalisation between arrays in LIMMA prior to further analyses.

The *lmFit* function of LIMMA was used to fit a linear model to the data. To take advantage of the within-array duplicate spots, the *duplicateCorrelation* function of LIMMA was implemented prior to the *lmFit* function in order to estimate the correlation between duplicate spots. This inter-duplicate correlation was then used as an argument in *lmFit*. Using the mother line at 8 dap as the reference sample, a design matrix and a contrast matrix was constructed and used in the *eBayes* function in order to produce gene lists ranked in order of evidence for differential expression. One such list was obtained for each contrast specified by the contrast matrix. This ranking does not give a threshold cut-off level for differential expression. However, the *classifyTestF* function of LIMMA, used with the *eBayes* values, can provide estimates of statistically significant differences in gene expression for the individual contrasts. Thus, this function was implemented on the obtained data. The estimated log₂ fold changes and log odds values for

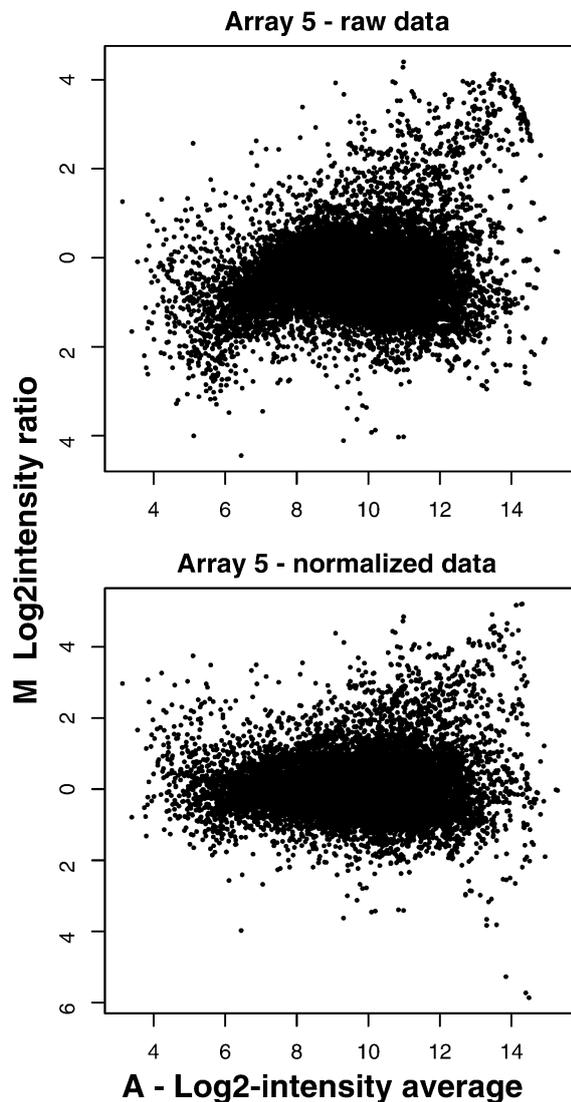


Figure 4. MA-plots for array 5 (Bobwhite, 16 vs. 8 dap contrast) prior to (a) and following (b) print-tip-loess normalisation with background subtraction and omission of control spots. M is \log_2 of the R/G signal ratio. A is \log_2 of average R and G spot signal intensities.

differential expression produced by the eBayes function were used to make visual presentations of the data, so-called volcano plots (Figures 6 and 7). The idea of the volcano plots is to reveal possible tendencies in the data, based on clouds of spots breaking away from a symmetrically ordered plot. So-called toptables, with arbitrarily chosen cut-off levels, were produced following the eBayes call, showing the identity of genes most likely to be differentially expressed in contrasts of interest and

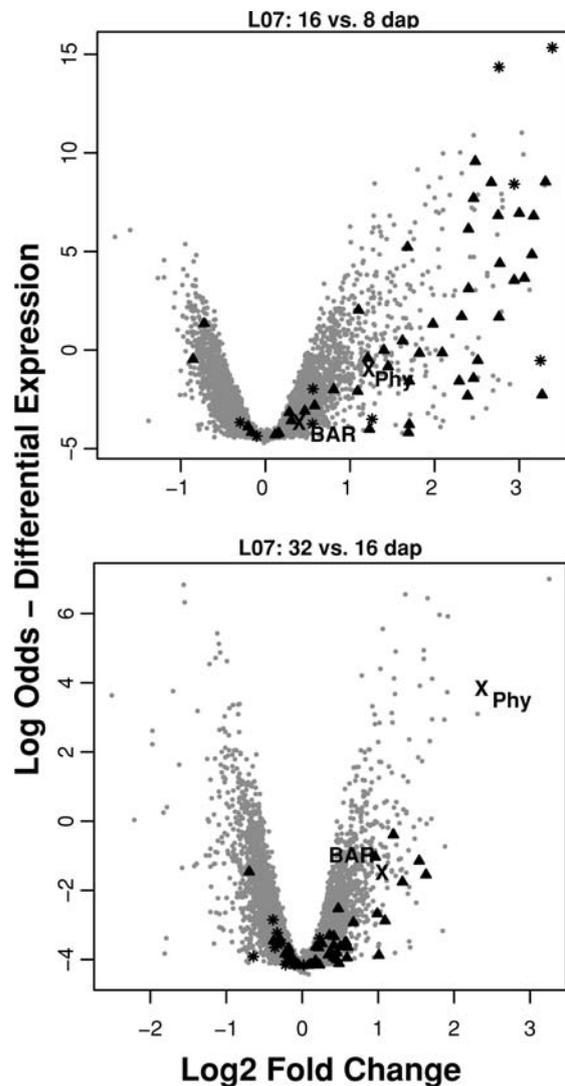


Figure 5. Changes in gene expression in wheat transgenic line L07 over three stages of the developing grain. Each point represents one gene in the 9K-wheat cDNA microarray. The log odds for differential expression of all genes, estimated from the LIMMA analysis of the data with a prior setting of differentially expressed genes of 5%, were plotted against the estimated \log_2 fold changes in the contrasts shown. *Black triangles*: Genes with annotations for the storage proteins gliadins and glutenins. *Asterisks*: genes with annotations for α -amylase inhibitors. BAR: the *bar* selection marker gene. Phy: the PhyFum-mod gene.

the accompanying p -values adjusted for multiple testing (Tables 1 and 2).

Differential expression over time

The experimental design (Figure 3) with developmental stage as one of the factors allowed for an

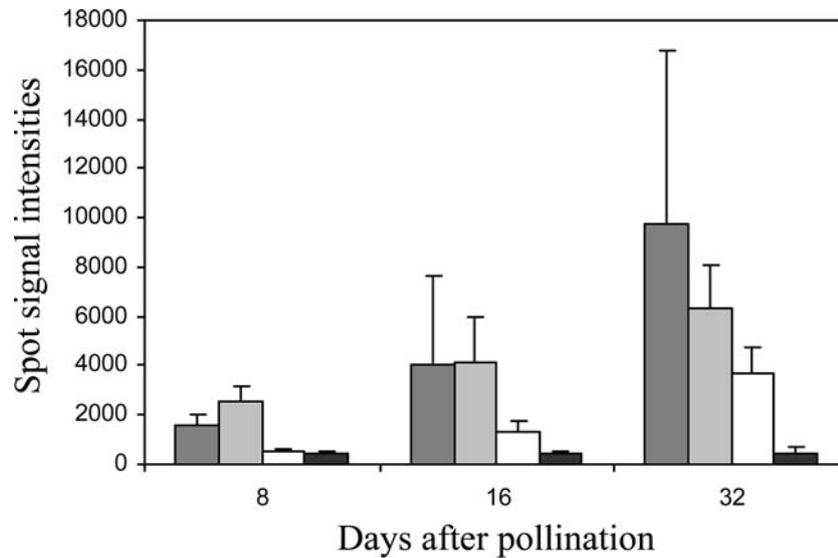


Figure 6. Quasi-quantitative gene expression levels in developing wheat seeds for HMW glutenin genes and the PhyFum-mod transgene driven by the 1DX5 HMW glutenin promoter. Following filtering for bad spots and PCR products, four clones with HMW glutenin annotations were included. The intensity signals are single channel values normalised between arrays according to the LIMMA User's Guide (www.bioconductor.org). Dark grey: average of four HMW glutenin clones, L07 samples; light grey: 1DX5 HMW glutenin clone, L07 samples; white: PhyFum-mod, L07 samples; black: PhyFum-mod, Bobwhite samples.

assessment of the potential differences between the two lines against the background of changes in gene expression levels over time. Results for the time contrasts are shown in Figure 6 and Table 1 for line L07. The results for the time contrasts for the mother line were almost identical to the L07 results, e.g. with regard to the identities of clones in the toptables (data not shown). For the calculations of the lods values used to create the volcano plots Limma requires a prior setting of the proportion of expected differentially expressed genes. In Figure 6 a prior setting of 5% was used since we expected a substantial number of genes to change expression over time. Especially between 8 and 16 dap annotations for well-known endosperm-accumulating proteins dominated the group of genes with changed, upregulated, expression levels, showing more than 8-fold increase for some of the genes. This is evident from the highlighting of genes with annotations for glutenins, gliadins, and α -amylase inhibitors in Figure 6a, where these genes dominate the divergent part of the spot cloud with positive M-values in the volcano plot. The eBayes call showed 412 genes with FDR-adjusted p -values ≤ 0.05 and of these the classifyTestsF function of Limma rated 80 genes as differentially expressed in this contrast. Table 1a

shows the identity of the top 50 genes most likely to be differentially expressed in this contrast, and also here it is evident that genes for seed storage proteins are prevalent. In addition, taken that there was no difference between expression patterns of the two lines, a combined analysis using the ClassifyTestF of Limma selected up to 370 genes from the time contrast 16 vs. 8 dap as differentially expressed (data not shown). Between 16 and 32 dap the expression levels of the group of seed storage protein genes seemed to have reached a maximum, and in some cases such as for some of the α -amylase inhibitors, even to have decreased. In this contrast, 70 genes had FDR-adjusted p -values ≤ 0.05 with a mixture of negative and positive M-values, and of these the classifyTestF function rated 37 genes as differentially expressed. Table 1b shows the identity of the top 30 genes of the Limma toptable from this contrast. Again, taken that there was no difference between expression patterns of the two lines apart from the transgenes, a combined analysis using the ClassifyTestF of Limma selected up to 220 genes from the time contrast 32 vs. 16 dap as differentially expressed (data not shown).

The phytase transgene, PhyFum-mod, and the selection gene, *bar*, showed increased expression

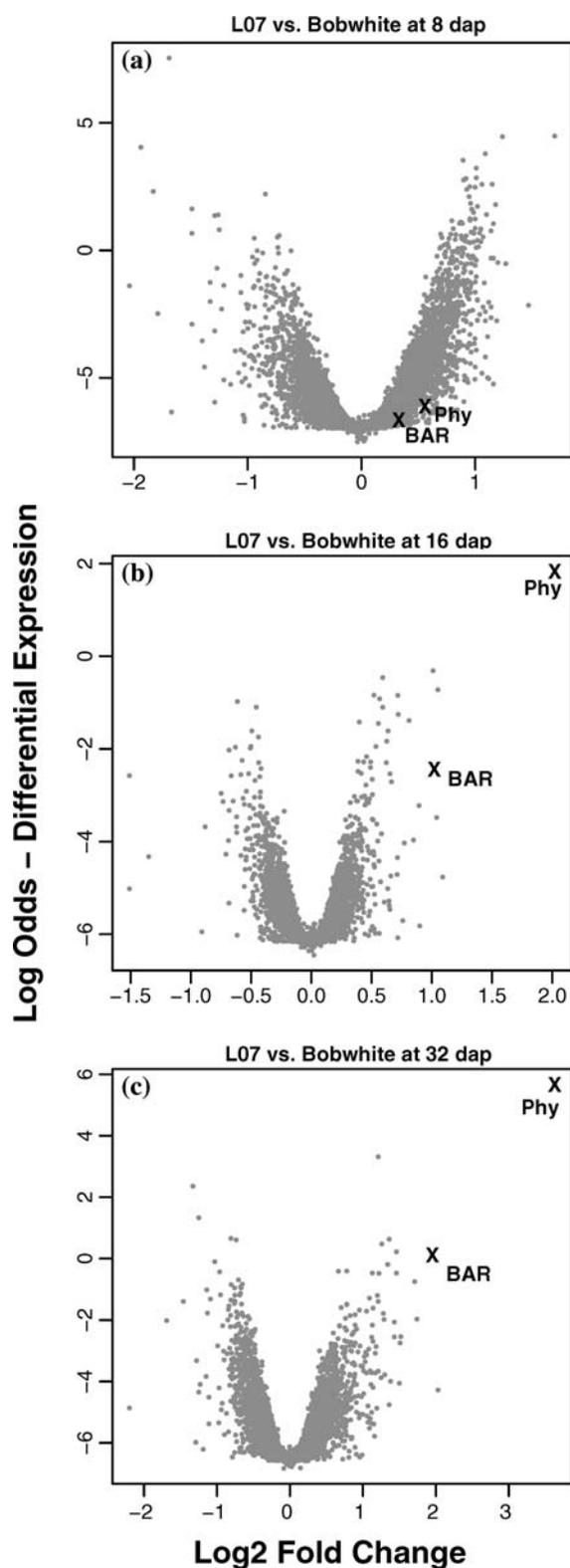


Figure 7. Comparisons of gene expression in the wheat transgenic line L07 and the control line Bobwhite at three stages (8, 16, and 32 days after pollination) of the developing grain. Each point represents one gene in the 9K-wheat cDNA microarray. The log odds for differential expression of all genes, estimated from the LIMMA analysis of the data with a prior setting of differentially expressed genes of 0.5%, are plotted against the estimated \log_2 fold changes in the contrasts shown. One outlier at $(-3.1, -3.3)$ was left out from B in order to improve visual quality of the presentation. BAR: the *bar* selection marker gene. Phy: the PhyFum-mod gene.

levels over time in line L07 as highlighted in Figure 6. In this case the contrasts are between samples of transgenic materials so the reservations noted above about estimation of transgene expression levels against a control are not relevant. The major increase in expression level of PhyFum-mod from 16 to 32 dap brought this gene into the group of significantly changed genes according to the classifyTestF procedure of LIMMA (Table 1b). An interesting aspect here is whether the phytase transgene that is driven by the 1DX5 promoter, showed an expression pattern and expression levels similar to the endogenous HMW glutenin gene driven by this promoter. To illustrate this, absolute expression values for the phytase transgene and the 1DX5 HMWglutenin, plus three other HMW glutenins with high-quality spots and PCR products, were extracted from the raw data using normalisation of single channel signals according to the LIMMA User's Guide. Figure 5 shows the expression profiles for line L07 for the HMW glutenins (1DX5 HMW glutenin and an average of the four HMW glutenins) and the PhyFum-mod transgene. For the latter the mother line values are included as well, in order to illustrate the non-specific hybridisation in the spots. The PhyFum-mod signal in line L07 at 6 dap was close to the non-specific hybridisation signal in the mother line. However, it increased steadily over time as also indicated in Figure 5. At 16 dap it was around 1/3 of the HMW glutenin level, and at 32 dap it was between 1/2 and 1/3 of the HMW glutenin levels. In the mother line the signal stayed at the same low level over the three time points.

Differential expression between lines

The differences in expression levels between line L07 and the mother line at the three different

Table 1. Top-ranked genes according to the log odds for differential expression in time contrasts for the transgenic wheat line L07: 16 dap vs. 8 dap, and 32 dap vs. 16 dap

Wheat EST clone name	Accession #	Log2 fold change	Fdr-adjusted <i>p</i> -value	Top blastx hit description
A. L07 – 16 vs.8 dap				
A04_h116_plate_9	AL815161	3.39	8.01E-08*	α -Amylase inhibitor CM3 precursor
D11_h116_plate_7	AL815061	2.76	2.23E-07*	α -Amylase inhibitor CM17 precursor
B09_d37_plate_02	–	3.03	2.17E-05*	α -2-purothionin precursor
B03_d37_plate_8	AL811443	2.46	2.17E-05*	Triticin precursor
C03_d37_plate_1	AL811130	2.30	4.53E-05*	Triticin precursor
E12_h116_plate_8	AL815136	2.10	4.53E-05*	None
A12_e310_plate_10	AL812043	3.05	4.53E-05*	None
C07_h116_plate_8	AL815121	2.48	6.21E-05*	LMW glutenin
G03_e310_plate_1	AL812273	1.80	9.32E-05*	MOTHER of FT and TF1 protein – <i>A. thaliana</i>
A06_h116_plate_13	AL814583	2.32	1.06E-04*	Diacylglycerol kinase – <i>A. thaliana</i>
C07_e310_plate_10	AL812055	2.07	1.31E-04*	Zinc finger protein – <i>A. thaliana</i>
B04_h116_plate_10	AL814385	3.31	1.38E-04*	Calcium-dependent protein kinase – <i>O. sativa</i>
F04_h116_plate_11	AL814486	2.67	1.38E-04*	B12D protein – <i>H. vulgare</i>
B05_p840_plate_3	AL829591	1.29	1.38E-04*	LMW glutenin
C10_h116_plate_13	AL814600	2.94	1.38E-04*	γ -Gliadin
F09_n129_plate_4	AL819757	1.93	1.38E-04*	α -Amylase inhibitor Ima1 precursor
C11_h116_plate_12	AL814537	3.30	1.38E-04*	RNA-binding protein-like – <i>A. thaliana</i>
E03_e310_plate_4	AL812482	2.79	2.08E-04*	Serpine
H03_d37_plate_1	AL811168	2.18	2.08E-04*	None
C12_N130_plate_48	AL821994	2.51	2.08E-04*	Lil3 protein – <i>A. thaliana</i>
E05_d37_plate_1	AL811144	2.00	2.28E-04*	Microtubule-binding protein TANGLED1
C06_h116_plate_7	AL815046	2.46	2.28E-04*	α/β -Gliadin A-IV precursor
B05_d37_plate_5	AL811352	2.80	2.48E-04*	None
H08_n129_plate_18	AL819337	1.89	3.07E-04*	None
A01_p840_plate_1	AL829484	2.80	3.38E-04*	18S rRNA
G12_h116_plate_8	AL815148	2.44	3.38E-04*	Orthophosphate dikinase
A11_h116_plate_6	AL814966	2.47	3.63E-04*	Trypsin inhibitor CMx precursor
A01_n129_plate_5	AL819772	2.10	3.63E-04*	PE-PGRS family protein
A09_d37_plate_1	AL811115	2.02	3.93E-04*	Polygalacturonase
H07_h116_plate_11	AL814507	2.05	3.99E-04*	None
C10_d37_plate_9	AL811497	3.00	3.99E-04*	α -Gliadin
B11_p234_plate_11	–	1.28	4.18E-04*	None
E01_e310_plate_11	AL812123	2.75	4.18E-04*	Seed storage protein
G03_N130_plate_1	AL820192	1.47	4.18E-04*	Cytochrome c oxidase subunit VIa precursor
H01_h116_plate_10	AL814428	3.17	4.22E-04*	α -Gliadin
F10_n129_plate_6	AL818956	1.77	5.92E-04*	GDA2 protein – <i>P. sativum</i>
D05_n129_plate_18	AL819297	1.52	6.66E-04*	Ferredoxin III
D09_e411_plate_14	AL813247	1.00	7.04E-04*	Transcriptional coactivator-like protein
D09_l226_plate_12	AL817852	1.34	7.04E-04*	OSK4
F06_e310_plate_10	AL812074	1.28	7.33E-04*	Catalase isozyme
E04_h116_plate_11	AL814475	2.09	7.48E-04*	RNA recognition motif (RRM)-containing protein-like
B03_h116_plate_1	AL814654	2.40	7.48E-04*	Secalin precursor
A12_l226_plate_2	AL816699	-1.60	7.79E-04*	Dehydrin
D12_n129_plate_6	AL819885	1.71	8.18E-04*	18S rRNA gene
F07_n129_plate_18	AL819319	1.36	8.18E-04*	myb protein family

Table 1. Continued

Wheat EST clone name	Accession #	Log2 fold change	Fdr-adjusted <i>p</i> -value	Top blastx hit description
C05_N130_plate_52	AL821910	1.26	8.39E-04*	Cytochrome P450
B04_h116_plate_11	AL814447	2.74	9.46E-04*	Extensin-like protein
C01_j223_plate_7	AL815582	1.37	1.01E-03*	Peroxiredoxin
G06_d26_plate_14	AL810956	-1.78	1.03E-03*	MtN3 protein precursor
D09_j324_plate_7	AL817267	1.43	1.12E-03*	Caleosin 1
B. L07 – 32 vs. 16 dap				
B02_l226_plate_07	AL817504	3.25	2.43E-03*	γ -1 purothionin
A02_h116_plate_08	AL815098	-1.56	2.43E-03*	B1108H10.20 – <i>O. sativa</i>
A11_l226_plate_03	AL816758	1.36	2.43E-03*	r40g2 protein – <i>O. sativa</i>
D09_j324_plate_07	AL817267	1.65	2.43E-03*	Abscisic acid-induced protein – <i>O. sativa</i>
C05_e29_plate_11	AL813580	-1.55	2.43E-03*	SERK2 protein – <i>Z. mays</i>
D12_n129_plate_06	AL819885	1.81	2.94E-03	None
G06_j223_plate_12	AL818153	1.92	2.94E-03*	PHD-finger family homeodomain protein – <i>O. sativa</i>
D05_j324_plate_07	AL817265	1.06	4.20E-03*	Globulin – <i>O. sativa</i>
B04_e29_plate_04	AL810542	-1.12	4.47E-03*	P0413C03.14 – <i>O. sativa</i>
E07_h116_plate_07	AL815068	-1.10	5.91E-03*	Cellulase – sweet orange
F10_n129_plate_06	AL818956	1.60	6.41E-03*	GDA2 protein – <i>P. sativum</i>
F07_l125_plate_09	AL818940	1.23	6.41E-03*	Diaminopimelate decarboxylase – <i>A. thaliana</i>
B04_p638_plate_05	AL827258	-1.08	6.41E-03*	None
E02_j324_plate_07	AL817270	-1.14	6.88E-03	None
D07_j223_plate_14	AL818261	1.60	6.88E-03*	Globulin Beg1 precursor – <i>H. vulgare</i>
C06_j324_plate_03	AL817108	-0.99	7.04E-03*	MYC transcription factor
E10_j223_plate_03	AL815337	-1.22	7.45E-03	None
B05_e411_plate_07	–	1.03	8.12E-03*	AMP deaminase
F12_p537_plate_08	AL826755	0.78	1.02E-02*	peroxidase – <i>O. sativa</i>
B06_n130_plate_15	AL821798	1.21	1.02E-02	T1N15.2 – <i>A. thaliana</i>
B08_p133_plate_14	AL821565	1.71	1.02E-02	C-4 sterol methyl oxidase – <i>O. sativa</i>
D07_n129_plate_06	AL819880	1.55	1.26E-02*	None
–	AY582135	2.36	1.32E-02*	PhyFum-mod
B10_a22_plate_10	AL808497	-1.70	1.41E-02*	None
G11_o232_plate_02	AL821093	1.91	1.42E-02	γ -2 purothionin
C03_l226_plate_11	AL817789	1.22	1.42E-02	Rab28 protein – <i>Z. mays</i>
D09_e29_plate_02	AL810427	-2.51	1.47E-02*	None
C03_p234_plate_02	AL824761	-0.82	1.85E-02*	Soluble starch synthase-like protein
A05_e29_plate_05	AL810605	-0.85	1.85E-02	Cell wall invertase
A03_p335_plate_10	AL822558	-0.91	1.85E-02	Chlorophyll a/b-binding protein

Cut-off levels of 50 and 30 genes were chosen for the two contrasts, respectively. *P*-values were adjusted for multiple testing using the false discory rate method of Limma.

*Genes rated as differentially expressed according to ClassifyTestF procedure of the LIMMA package.

harvest time points are shown in the volcano plots of Figure 7 and in the toptables of Table 2. For the creation of volcano plots, a prior setting of 0.5% differentially expressed genes was used for the calculation of the lods values in Limma, since we expected only slight differences between the two lines. At 8 dap, several genes were classified as

having different expression levels between the two lines, with both up- and down-regulation of quite a number of genes in the transgenic line, however, with a moderate range of M-values from -2 to 2 and a symmetric volcano plot without clear tendencies in the spot cloud. We deciphered the expression pattern for 36 endosperm-expressed

Table 2. Top-ranked genes according to the log odds for differential expression in the contrasts between the wheat line L07 and cv. Bobwhite at three stages of the wheat grain development: 8, 16, and 32 days after pollination

Wheat EST clone name	Accession #	Log2 fold change	Fdr-adjusted <i>p</i> -value	Top blastx hit description
A. L07 vs. Bobwhite at 8 dap:				
D11_h116_plate_7	AL815061	-1.69	0.0005*	α -Amylase inhibitor, CM17 precursor
A12_l226_plate_2	AL816699	1.70	0.0060*	Dehydrin – durum wheat
B01_j324_plate_8	AL817302	1.24	0.0060*	None
B09_d37_plate_2	–	-1.94	0.0071*	α -2-purothionin precursor
G04_j223_plate_7	AL815619	1.09	0.0075	None
B10_p335_plate_19	AL825571	0.89	0.0082*	HGA6 – <i>H. vulgare</i>
C12_a11_plate_4	AL809524	1.01	0.0099	Trehalase – <i>O. sativa</i>
D03_e310_plate_8	AL812745	1.01	0.0110	None
D12_d37_plate_5	AL811363	0.92	0.0110*	Importin β -like protein
A04_p537_plate_4	AL826516	0.90	0.0110*	GAD1 – <i>H. vulgare</i>
B09_j324_plate_4	AL817155	1.15	0.0110*	Repair endonuclease – <i>A. thaliana</i>
H07_p234_plate_27	AL822553	1.06	0.0110	Protein phosphatase 2C – <i>O. sativa</i>
C06_j324_plate_3	AL817108	0.95	0.0110*	MYC transcription factor – <i>O. sativa</i>
B04_e29_plate_4	AL810542	1.00	0.0110*	P0413C03.14 – <i>O. sativa</i>
E06_p638_plate_8	AL827018	0.93	0.0113	Adenosine kinase – <i>O. sativa</i>
B. L07 vs. Bobwhite at 16 dap:				
–	AY582135	2.02	0.0443*	PhyFum-mod
B05_e310_plate_2	AL812296	1.01	0.4369	Protein disulfide isomerase
H10_o232_plate_2	AL821148	0.59	0.4369	60S ribosomal protein L41 – <i>O. sativa</i>
C06_h116_plate_7	AL815046	1.05	0.4369	α/β -Gliadin MM1 precursor
E11_p840_plate_4	AL829651	0.52	0.4369	None
C07_e29_plate_11	AL813581	0.72	0.4369	Protein phosphatase type-2C – <i>Z. mays</i>
H06_o232_plate_2	AL821124	0.57	0.4369	Thiamine biosynthesis protein ThiC – <i>P. secunda</i>
F07_l125_plate_9	AL818940	-0.61	0.4369	Diaminopimelate decarboxylase – <i>A. thaliana</i>
D07_p234_plate_12	AL825058	-0.46	0.4369	None
G02_p638_plate_3	AL827359	0.59	0.4369	LRR receptor-like kinase 2 – <i>O. sativa</i>
C02_p335_plate_9	AL825373	0.72	0.4815	ABA-responsive protein – <i>H. vulgare</i>
B01_e512_plate_9	AL814361	0.81	0.4815	RuBisCO
F09_e411_plate_14	AL813257	0.40	0.4815	Proline-rich protein
A12_n130_plate_39	AL821736	0.56	0.4815	Glycine-rich RNA-binding protein – <i>H. vulgare</i>
B12_d26_plate_16	AL811022	-0.49	0.4815	ESR2g1 – <i>Z. mays</i>
C. L07 vs. Bobwhite at 32 dap:				
–	AY582135	3.63	0.0015*	PhyFum-mod
D11_h116_plate_7	AL815061	1.21	0.0183*	α -Amylase inhibitor, CM17 precursor
H12_n130_plate_1	AL820407	-1.33	0.0272*	None
F10_p436_plate_12	AL828225	-1.25	0.1049	Acyl-CoA synthetase – <i>O. sativa</i>
E10_p436_plate_10	AL828107	-0.81	0.1422	DNA binding protein RAV2 – <i>O. sativa</i>
F06_e29_plate_5	AL810636	1.36	0.1422*	γ -Gliadin
D07_p234_plate_12	AL825058	-0.74	0.1422	None
B12_e512_plate_8	AL814317	1.26	0.1451	Secretory acid phosphatase precursor – <i>O. sativa</i>
A02_p739_plate_3	AL827892	1.46	0.1727	None
–	X05822	1.95	0.1754	BAR
G02_p234_plate_6	AL824902	-1.03	0.1998	Protein sam2B – spinach
G11_o232_plate_15	AL820630	1.34	0.1998	β -Thionin precursor – <i>H. vulgare</i>
D12_p335_plate_18	AL825534	0.78	0.1998	RuBisCO, small subunit
C12_o232_plate_13	AL820524	0.66	0.1998	None
D11_p436_plate_7	AL826215	-0.96	0.1998	None

A cut-off level of 15 genes was chosen. *P*-values were adjusted for multiple testing using the false discovery rate method of Limma.

*Genes rated as differentially expressed according to ClassifyTestF procedure of the LIMMA package.

glutenin and gliadin genes in order to evaluate if the differential expression was due to presence of the transgene or resulted from asynchronous development of the seed in the two lines. For these 36 genes there seemed to be a bias towards higher expression in the mother line, inferred from a negative mean M-value of -0.22 for the L07 vs. the mother line contrast. Two genes encoding an α -amylase inhibitor and an α -2-purothionin showed large negative M-values in the top 15 toptable (Table 2A). These proteins accumulate along with the main storage proteins of the wheat seed (e.g. Shewry et al., 2002) and the differences in expression levels may accordingly also result from asynchrony in grain development. As indicated in Figure 7, the expression level of the phytase was very low at the early stage, and in concordance with that it did not show up in the toptable as differentially expressed.

At 16 dap the phytase transgene was highly expressed and was the only gene selected as differentially expressed by classifyTestF and the only gene with a FDR-adjusted p -value < 0.05 , as shown in the top 15 toptable of Table 2B. The expression of the 36 glutenin and gliadin genes was very similar in the two types of material, with a mean M-value of -0.02 , indicating well-synchron-

ised development of the plant material going into the analysis at this stage.

Also at 32 dap the phytase transgene showed high expression. The other transgene in line L07, the *bar* selection gene, was also included in the top 15 toptable of Table 2C, although it was not rated as differentially expressed by the classifyTestF method of LIMMA, presumably due to low expression levels. A few other genes were rated as differentially expressed by classifyTestF, the top scorer being an α -amylase inhibitor gene. This gene was also one of the only two genes in addition to PhyFum-mod that had FDR-adjusted p -values ≤ 0.05 . Also for this contrast the over-all pattern of the 36 glutenin and gliadin genes was investigated, and the mean M-value of this group of genes was 0.33 for this contrast. This indicated a bias towards higher expression of storage protein genes in L07 compared to Bobwhite and, thus, probably slight asynchronous development of the plant material at this harvesting time point.

Real time PCR validations

Real time PCR was used to validate the microarray results for the late time point (Table 2C) where a few more genes than PhyFum-mod

Table 3. Results from real time PCR validation of the differential expression between the transgenic line L07 and the control line, Bobwhite, for the seven genes following the PhyFum-mod gene in the toptable of the 32 dap contrast (Table 2C)

Genes acc. #	M (log2 fold change: L07 vs. Bobwhite, 32 dap)				
	Means of all three samples per line	Sample pair used for microarray 4	Sample pair not used in microarray study	Estimated microarray M-values, contrast L07-control, 32 dap	Rel. expression levels 32 dap: Average spot intensities in array 4 relative to background
α -Amylase inhibitor AL815061	0.45	0.66	-0.08	1.21	65
(Unknown) AL820407	0.11	0.03	0.65	-1.33	11
Acyl-CoA synthetase AL828225	-0.21	-0.67	-0.24	-1.25	5
DNA binding protein AL828107	0.41	-1.25	0.54	-0.81	4
γ -Gliadin AL810636	0.96	1.04	0.38	1.36	18
(Unknown) AL825058	0.07	-0.46	-0.10	-0.74	3
Acid phosphatase AL814317	-0.01	-0.20	0.25	1.26	8

Three independent samples from each of the two lines were used, comprising the two samples per line used in hybridisations in the microarray study and one additional independent sample per line. The expression levels were normalised relative to 18S expression levels in the individual samples. The M values indicated are log2 ratios of expression levels between contrasting samples or means of samples. M-values were calculated as means of four to six PCR reactions per sample. For comparisons, the M-values from Table 2C are given, and also an estimate of the expression levels of the genes, based on the intensity levels in the array 4: L07 vs. Bobwhite, 32 dap.

indicated differential expression. Thus, additional seven genes from the toptable of L07-control contrast at 32 dap were selected for real time PCR analysis. In addition, real time PCR was performed using primers for the PhyFum-mod and the BAR genes, which positively confirmed their expression in the transgenic lines (data not shown). Three independent sample batches from both the transgenic and the mother line were subjected to real time PCR, comprising the two sample batches used in hybridisations of the microarray study and one additional, independent sample batch. Table 3 shows the log₂ fold differences (M-values) between the transgenic L07 line and the mother line obtained from the real time PCR analysis, either of the indicated contrasting samples or means of samples.

Taken as a whole, the real time PCR result could not confirm differential expression of the seven genes in the 32 dap contrast between the two lines. First, most of the calculated M-values were less extreme than the M-values from the microarrays, although the tendencies for the highly expressed seed storage protein genes, the α -amylase inhibitor and the γ -gliadin, seemed to confirm the bias towards expression of these genes in line L07 in the sample batches used for array 4. Second, the independent third pair of sample batches that was not used in the microarray study, seemed to have more balanced expression between the two lines for all the seven genes. Probably this reflects a better synchrony of the developmental processes in the harvested plant material for these two independent sample batches. The calculated M-values varied more for genes expressed at low levels, e.g. the acyl-CoA synthetase and the DNA binding protein, and this presumably ensues intrinsic properties of the PCR process. The fact that these two genes were not rated as differentially expressed by classifyTestF probably reflects a higher signal variation also in the microarrays for these genes.

Discussion

In the present study we have performed a comparative analysis of the gene expression profiles during grain development in a transgenic wheat line and the mother variety using a 9K cDNA microarray. The primary objective of the analysis

was to evaluate microarray-based expression profiling as a tool for detecting potential unintended changes in gene expressions in the transgenic plants. In theory, transgene integration itself or effects on downstream processes by the transgene encoded protein might lead to a modification of cellular metabolism that potentially could result in the formation of compounds with adverse effects on human and animal health. The implementation of global profiling techniques that allow for analyses of the gene expression, protein, and secondary metabolite profiles in the transgenic plants compared to their mother varieties might be important tools for addressing these concerns.

To our knowledge, the current study is the first published attempt to evaluate microarray-based gene expression analyses to be used in safety assessments. We have presented a case where the wheat cultivar Bobwhite was genetically transformed with a phytase encoding gene from *A. fumigatus*. Codon modification of the gene towards a codon usage preferred by wheat ensured high levels of phytase activity in flour of mature wheat seeds. The gene was designed for endosperm-specific expression by using the wheat 1DX5 HMW glutenin promoter and for targeting to the apoplast via the insertion of a signal sequence derived from an α -amylase gene of barley. For the microarray studies we selected a line, L07, with a high phytase activity level (average of 4571 FTU/kg) and a single integration site for the phytase transgene. The phytase activity level of line L07 was in the range of what was previously obtained after endosperm specific (glutelin promoter) expression of the *A. fumigatus* gene in rice and was higher than the activity levels previously achieved in wheat using a native *Aspergillus niger* phytase open reading frame with the maize ubiquitin promoter, and a shorter (440 bp) version of the 1DX5 promoter (Brinch-Pedersen et al., 2000, 2003; Lucca et al., 2001). Metal dye detection HPLC analyses of the latter material did not reveal any differences in the myoinositol phosphate profile in the mature seeds of wild type and the transgenic line strongly indicating that the expressed fungal phytase is not acting against phytate during development of the grain, presumably due to compartmentalisation (Brinch-Pedersen et al., 2002, 2003).

The experimental design for the microarray analyses comprised sampling of grain RNA from the transgenic line and the mother variety between

12 am and 2 pm at three different time points during grain development, specified as dap. This allowed for an analysis of differential expression between the two lines against the background of developmental changes in the wheat grain. Large-scale changes in gene expression are to be expected during grain filling, preceding drastic morphological and metabolic changes (for a review on grain development see e.g. Simmonds & O'Brien, 1981). A detailed study of the developmental changes in gene expression is outside the scope of this work. However, inspection of the data showed that, for example, genes for the storage proteins gliadins and glutenins were prevalent among the genes changing expression over time. We used the expression of these genes to compare the developmental status of contrasting samples. Comparisons of different harvesting time points evidently showed strong changes in expression of these genes and, on the other hand, slight differences in expression of these genes for material harvested at the same time point indicated asynchronous development of the plant material. The expression levels over time of the phytase gene, driven by the 1DX5 HMW glutenin promoter, appeared to parallel that of the HMW glutenins, as previously well-documented for this promoter (Shewry et al., 1994). Also the *bar* gene was properly identified in the transgenic line although with expression levels considerably lower than for the phytase gene.

The tests for differential gene expression between the two wheat lines showed slight variations for the three sampling time points. The phytase transgene was clearly "differentially" expressed at the intermediate and late time point. At the early time point, 8 dap, a number of genes were rated as differentially expressed, although with moderate *M*-values between -2 and 2 and with a symmetrically distributed volcano plot without any tendencies in the spot cloud. At the intermediate time point, 16 dap, only the transgene itself was selected as differentially expressed. At 32 dap the transgene topped the table of differentially expressed genes with a few other genes, though, rated also as differentially expressed by classify-TestF of Limma. However, results of real-time PCR for the seven genes following the phytase transgene in the toptable for the 32 dap contrast between the two lines could not confirm differential expression of these genes. Furthermore, inspection of the data regarding expression profiles

of storage protein genes revealed some biases between the two lines at the early and the late time points. We thus interpret the differential expressions of other genes than the transgene at the early and the late time point to be a reflection of asynchrony in development between the two materials. The general conclusion from the present study therefore is that we have found no conclusive evidence for significant changes in gene expression profiles between the transgenic line and the mother variety.

The slight variations in the results underline that proper sampling techniques are of paramount importance in DNA microarray experiments, and that in our case the number of replicates used in the experimental design was presumably not high enough to counter the effects of asynchronous grain development. Furthermore, our results also stress that careful interpretation of microarray results are necessary when studying differences in a biological background of extensive progressive changes like the developing seed (Kuiper et al., 2003). Thus, genes showing large changes in expression during a relatively short period can erroneously be identified as differentially expressed. This was probably the case for the samples taken 8 dap, since this is a stage when large developmental changes take place in the seed (see e.g. Simmonds & O'Brien, 1981). A recent paper (Leader, 2005), presenting microarray analyses in wheat similar to ours, showed an upsurge in the amounts of e.g. glutenin gene transcripts around 8 dap. Thus, the chance of harvesting material out of synchrony is probably high at this stage. With these reservations about the early sampling time point, we believe that the more robust results of the intermediate and late sampling time points support the conclusion that there is no evidence for significant changes in gene expression profiles between the transgenic line and the mother variety. Following completion of this work, a recent paper (Ouakfaoui & Miki, 2005) reported, that the integration and expression of two selection marker genes, *nptII* and *qusA*, in *A. thaliana* did not give rise to overall changes in gene expression patterns.

Two inherent shortcomings of the present study should be taken into consideration. The used cDNA microarray contains 9000 unigenes which is only a fraction of the wheat gene complement. On the other hand, the unigene set

is primarily based on cDNA libraries generated from the developing wheat grain. Recent estimates suggest that the wheat endosperm transcriptome only comprises some 4500–8000 genes (Clarke et al., 2000). Thus, the 9K wheat microarray probably gives a rather good representation of genes expressed in the developing grain. The second shortcoming relates to the specificity of the array and, in general, to the sensitivity of the cDNA microarray technology. Some cross hybridisations will occur when using cDNA microarrays and much higher specificity, and perhaps also sensitivity, could be achieved by oligo-based arrays. However, our conclusion is that while the experimental setup is not ideal it does provide a reasonably robust procedure for detection of major changes in the transcriptome in the transgenic line used in this study. An absolute verification of the absence of unintended effect might require additional approaches such as back crosses to the mother variety to remove potential somaclonal variation (for discussion, see Cellini et al., 2004), detailed studies of the transgene integration sites, and information on the localisation and *in planta* action of the heterologous enzyme.

Ideally, an assessment of novel types of plants should comprise global analyses of the transcriptome, the proteome and the metabolome. While large gene arrays are rapidly becoming available for several crop plants it will take several years before similar tools are available for describing the cell constituents and primary and secondary metabolites of plants (see discussion by Cellini et al., 2004). However, we believe that the present work has shown that the transcript profiling technique is feasible in a safety assessment context, and that in the future this technique, together with other profiling techniques, could play an important role in the characterisation and safety assessment of transgenic crop plants.

Acknowledgements

We wish to thank Ian Wilson and Keith Edwards, Bristol University, for very kindly making their 9K wheat unigene collection available to us; Per Horn and Jakob Hedegård from the Danish Institute of Agricultural Sciences, Dept. of Animal Breeding and Genetics for helping with the spotting of microarrays; Lis Bagenkop Holte and

Ole Bråd Hansen for skilful technical assistance and for taking care of plants; and Gordon Smyth, WEHI, Australia, for very helpful assistance in the statistical data analysis. This work was supported by the Danish Ministry of Food, Agriculture and Fisheries, under the research programme Biotechnology in Food Research (BIOT99-13).

References

- Alba R, Fei Z, Payton P, Liu Y, Moore SL, Debbie P, Cohn J, D'Ascenzo M, Gordon JS, Rose JKC, Martin G, Tanksley SD, Bouzayen M, Jahn MM and Giovannoni J (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. *Plant J* **39**: 697–714.
- Brinch-Pedersen H, Galili G, Knudsen S and Holm PB (1996) Engineering of the aspartate family biosynthetic pathway in barley (*Hordeum vulgare* L.) by transformation with heterologous genes encoding feed-back-insensitive aspartate kinase and dihydropicolinate synthase. *Plant Mol Biol* **32**: 611–620.
- Brinch-Pedersen H, Olesen A, Rasmussen SK and Holm PB (2000) Generation of transgenic wheat (*Triticum aestivum* L.) for constitutive accumulation of an *Aspergillus phytase*. *Mol Breed* **6**: 195–206.
- Brinch-Pedersen H, Sørensen LD and Holm PB (2002) Engineering crop plants: getting a handle on phosphate. *Trends Plant Sci* **7**: 118–125.
- Brinch-Pedersen H, Hatzack F, Sorensen LD and Holm PB (2003) Concerted action of endogenous and heterologous phytase on phytic acid degradation in seed of transgenic wheat (*Triticum aestivum* L.). *Transgenic Res* **12**: 649–659.
- Cellini F, Chesson A, Colquhoun I, Constable A, Davies HV, Engel KH, Gatehouse AM, Kärenlampi S, Kok EJ, Leguay JJ, Lehesranta S, Noteborn HP, Pedersen J and Smith M (2004) Unintended effects and their detection in genetically modified crops. *Food Chem Toxicol* **42**: 1089–1125.
- Christensen AH, Sharrock RA and Quail PH (1992) Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Mol Biol* **18**: 675–689.
- Clarke BC, Hobbs M, Skylas D and Appels R (2000) Genes active in developing wheat endosperm. *Funct Integr Genomics* **1**: 44–55.
- Dudoit S, Yang YH, Callow MJ and Speed TP (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Stat Sinica* **12**: 111–140.
- Eisen MM and Brown PO (1999) DNA arrays for analysis of gene expression. *Method Enzymol* **303**: 179–205.
- Fiehn O, Kloska S and Altman T (2001) Integrated studies on plant biology using multiparallel techniques. *Curr Opin Biotechnol* **12**: 82–86.
- Kok EJ and Kuiper HA (2003) Comparative safety assessment for biotech crops. *Trends Biotechnol* **21**: 439–444.
- Kuiper HA, Kok EJ and Engel K-H (2003) Exploitation of molecular profiling techniques for GM food safety assessment. *Curr Opin Biotechnol* **14**: 238–243.

- Leader DJ (2005) Transcriptional analysis and functional genomics in wheat. *J Cereal Sci* **41**: 149–163.
- Lucca P, Hurrell R and Potrykus I (2001) Genetic engineering approaches to improve the bioavailability and the level of iron in rice grains. *Theor Appl Genet* **102**: 392–397.
- Ouakfaoui SE and Miki B (2005) The stability of the Arabidopsis transcriptome in transgenic plants expressing the marker genes *nptII* and *uidA*. *Plant J* **41**: 791–800.
- Pasamontes L, Haiker M, Wyss M, Tessier M and van Loon APGM (1997) Gene cloning, purification, and characterization of a heat-stable phytase from fungus *Aspergillus fumigatus*. *Appl Environ Microbiol* **63**: 1696–1700.
- Quackenbush J (2002) Microarray data normalization and transformation. *Nat Biotechnol Suppl* **22**: 496–501.
- Rogers JC and Milliman C (1983) Isolation and sequence analysis of a barley α -amylase cDNA clone. *J Biol Chem* **258**: 8169–8174.
- Shewry PR, Beaudoin F, Jenkins J, Griffiths-Jones S and Mills ENC (2002) Plant protein families and their relationships to food allergy. *Biochem Soc Trans* **30**: 906–910.
- Shewry PR, Tatham AS, Halford NG, Barker JHA, Hannappel U, Gallois P, Thomas M and Kreis M (1994) Opportunities for manipulating the seed protein composition of wheat and barley in order to improve quality. *Transgenic Res* **3**: 3–12.
- Simmonds DH and O'Brien TP (1981) Morphological and biochemical development of the wheat endosperm. In: Pomeranz Y (ed), *Advances in Cereal Science and Technology*. (pp. 6–70) IV American Society of Cereal Chemists, Inc, St. Paul, Minnesota.
- Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**(1).
- Smyth GK and Speed TP (2003) Normalization of cDNA microarray data. *Methods* **31**: 265–273.
- Smyth GK, Yang Y-H and Speed TP (2003) Statistical issues in microarray data analysis. In: Brownstein MJ and Khodursky AB (eds), *Functional Genomics: Methods and Protocols, Methods in Molecular Biology*. (pp. 111–136) Humana Press, Totowa, NJ.
- Thompson CJ, Novva NR, Tizard R, Cramer R, Davies JE, Lauwereys M and Botterman J (1987) Characterization of the herbicide resistance gene *bar* from *Streptomyces hygroscopicus*. *EMBO J* **6**: 2519–2523.
- Trethewey RN (2004) Metabolite profiling as an aid to metabolic engineering in plants. *Curr Opin Plant Biol* **7**: 196–201.
- Wilson ID, Barker GLA, Beswick RW, Shepherd SK, Lu C, Coghill JA, Edwards D, Owen P, Lyons R, Parker JS, Lenton JR, Holdsworth MJ, Shewry PR and Edwards KJ (2004) A transcriptomics resource for wheat functional genomics. *Plant Biotechnol J* **2**: 495–506.
- Yang HY, Dudoit S, Luu P, Lin DM, Peng V, Ngai J and Speed TP (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30**: e15.