

# Transgenesis has less impact on the transcriptome of wheat grain than conventional breeding

María Marcela Baudo<sup>1</sup>, Rebecca Lyons<sup>1</sup>, Stephen Powers<sup>1</sup>, Gabriela M. Pastori<sup>1,†</sup>, Keith J. Edwards<sup>2</sup>, Michael J. Holdsworth<sup>3</sup> and Peter R. Shewry<sup>1,\*</sup>

<sup>1</sup>Rothamsted Research, Harpenden AL5 2JQ, UK

<sup>2</sup>School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK

<sup>3</sup>Division of Agricultural and Environmental Sciences, School of Biosciences, University of Nottingham, Nottingham NG7 2RD, UK

Received 12 January 2006;

revised 7 March 2006;

accepted 13 March 2006

\*Correspondence (fax +44 (0)1582 763010; e-mail peter.shewry@bbsrc.ac.uk)

†Present address: BBSRC, Polaris House, North Star Avenue, Swindon SN2 1UH, UK

The gene expression data have been deposited in ArrayExpress database with the accession number A-MEXP-177; <http://www.ebi.ac.uk/arrayexpress/Submissions/index.html>

**Keywords:** transgenic wheat, transcriptome analysis, substantial equivalence.

## Summary

Detailed global gene expression profiles have been obtained for a series of transgenic and conventionally bred wheat lines expressing additional genes encoding HMW (high molecular weight) subunits of glutenin, a group of endosperm-specific seed storage proteins known to determine dough strength and therefore bread-making quality. Differences in endosperm and leaf transcriptome profiles between untransformed and derived transgenic lines were consistently extremely small, when analysing plants containing either transgenes only, or also marker genes. Differences observed in gene expression in the endosperm between conventionally bred material were much larger in comparison to differences between transgenic and untransformed lines exhibiting the same complements of gluten subunits. These results suggest that the presence of the transgenes did not significantly alter gene expression and that, at this level of investigation, transgenic plants could be considered substantially equivalent to untransformed parental lines.

## Introduction

Transgenesis offers immense opportunities for making a major contribution to the increased food requirement necessary to feed the expanding world population (estimated to reach 10 billion by the middle of this century, see Evans, 1993). There has been, and continues to be, considerable public resistance to the use of genetic modification (GM) technology for crop improvement, particularly in Western Europe and in relation to staple food crops such as wheat. This resistance stems from a number of different concerns that includes the perception that the apparently random insertion of transgenes into host plant genomes may result in unpredicted effects on the expression of other genes, leading to unforeseen effects on plant phenotype (e.g. increases in toxins and allergies). If this were to be the case, transgenic crops could not be considered to be 'substantially equivalent' to non-GM crops and may therefore require more extensive evaluation than conventionally bred lines prior to approval for food use (FAO/WHO, 1996).

The analysis of global gene expression changes at the level of RNA accumulation has been considerably enhanced by the development of microarrays, which allow the simultaneous analysis of the expression patterns of many thousands of genes. We recently reported the development and use of a cDNA-derived microarray for wheat, comprising a unique set of 10 000 elements, and described its use for the analysis of transcriptome changes associated with grain development and germination (Wilson *et al.*, 2004, 2005). The same array has also been used in a recent comparative study of gene expression in transgenic and nontransgenic wheat expressing a fungal phytase gene (Gregersen *et al.*, 2005).

In this paper we have used this microarray to address the issue of substantial equivalence at the level of RNA expression, by carrying out detailed transcriptome analyses of transgenic and conventionally bred wheat lines expressing the same wheat genes encoding (gluten) proteins under the control of their own endosperm-specific promoters, but with additional marker and reporter genes being present in the transgenic line. Furthermore, two additional transgenic lines

were generated in order to compare the transcriptome profiles of lines transformed with whole plasmids or with excised fragments of DNA containing only the gene of interest and the selectable marker gene ('clean genes'). Genes encoding gluten proteins were chosen for this analysis as it is considered that gluten proteins have a single role, as storage proteins, being deposited in the developing endosperm and mobilized during germination to support seedling growth. Therefore, altered expression of the genes encoding these proteins should not, in principle, influence the global expression patterns of genes in the developing endosperm, although some compensatory effects on the expression of related storage protein genes might be expected. Analysis of the results, incorporating strict statistical treatment of data, provides strong evidence that expression of the transgenes studied here has very little statistically significant impact on global genomic expression in the developing grain, particularly when compared to the large differences observed between sister lines produced by 'conventional breeding' approaches. Similarly, few statistically significant differences in genome expression were observed when the transformation was carried out using either whole plasmids or 'clean genes'.

This report forms part of a detailed series of studies of the substantial equivalence of the series of transgenic wheat lines reported by Barro *et al.* (1997). The accompanying paper (Baker *et al.*, 2006) reports detailed metabolomic studies of lines grown under field conditions (including L88-31 and B102-1-1, which are studied here) while the agronomic and functional properties of the same lines are reported by Darlington *et al.* (2003) and Shewry *et al.* (2006).

## Results

### Comparison of transgenic, control and conventionally bred lines

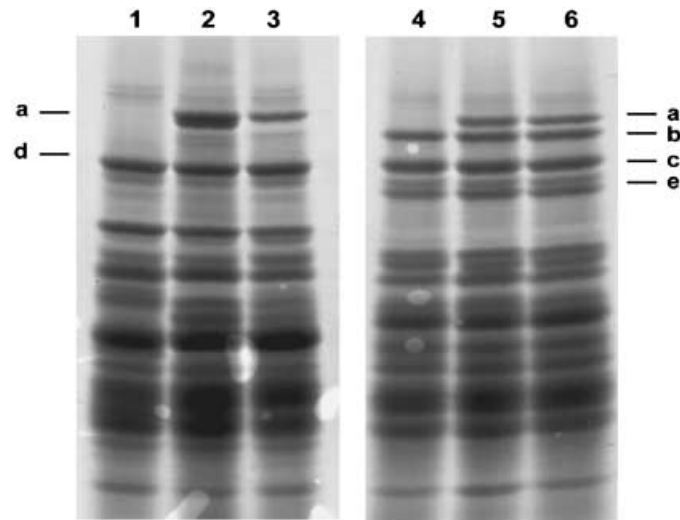
The high molecular weight (HMW) subunits of wheat glutenin are synthesized only in the starchy endosperm tissue of the developing wheat grain. They have been studied in detail because of their role in determining grain processing quality (Payne, 1987; Shewry *et al.*, 2003). All cultivars of hexaploid bread wheat contain six HMW glutenin subunit genes, two each at the *Glu-A1*, *Glu-B1* and *Glu-D1* loci on the long arms of the group 1 chromosomes (Payne, 1987; Shewry *et al.*, 2003). However, gene silencing results in only three, four or five subunits being expressed in commercial cultivars. In addition, null forms of all six genes can be identified in wider surveys of germplasm and can be used to generate series of related lines differing in their numbers of expressed HMW

glutenin subunit genes (Shewry *et al.*, 2003). One such series was selected for transformation, which was produced by crossing null lines in the Australian cultivars Olympic and Gabo (Lawrence *et al.*, 1988). One of these lines, L88-31, expresses only two HMW glutenin subunits encoded by *Glu-B1* (1Bx17, 1By18) while line L88-18 expresses the same two subunits in addition to subunit 1Ax1 encoded by *Glu-A1*. These lines were developed from a cross between the cultivar Gabo and a mutant line of cultivar Olympic which was null for the 1B-encoded subunits. The F<sub>1</sub> was selfed to give an F<sub>2</sub> family and the L88-18 and L88-31 lines selected from single F<sub>2</sub> seeds (G. Lawrence, CSIRO, Canberra, Australia, pers. commun.). The L88-18 and L88-31 lines therefore both have approximately equal contributions of genes from the two parents. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of seed proteins was used to confirm the homogeneity of the parental lines before transformation and of all of the lines at all stages of multiplication. However, some residual heterozygosity in the L88-31 and L88-18 lines cannot be ruled out. The L88 series of lines are similar in their degree of relatedness to many conventionally bred cultivars in which an F<sub>1</sub> is frequently crossed to a third variety, rather than backcrossed to one of the parents, in order to introduce specific traits and improve the overall performance at the same time (Dr W. Angus, Nickerson UK Ltd, Bury St. Edmunds, UK, pers. commun.).

The transgenic line B102-1-1 was produced by transforming L88-31 with the 1Ax1 gene (from wheat cultivar Hope) as a whole plasmid, and contains a single copy of the transgene integrated at the same locus as the pAHC25 plasmid which contains the *bar* selectable marker gene (Barro *et al.*, 1997; Rooke *et al.*, 2003). These three lines (L88-31, L88-18 and B102-1-1) therefore allow two comparisons to be made; between transgenic and untransformed lines, and between transgenic lines and sister lines conventionally bred to contain the same HMW subunit composition. SDS-PAGE analyses of the gluten proteins of the three lines are shown in Figure 1.

The effects of introducing transgenes on gene expression profiles in developing seeds of B102-1-1 and L88-31 were determined using total RNA prepared from endosperms at 14 and 28 days post-anthesis (dpa). Similar studies were also carried out on leaves at 8 days post-germination (dpg), a tissue in which the *bar* gene but not the *1Ax1* gene is expressed. Results are summarized in Table 1.

Microarray comparisons of the endosperm mRNA fractions at 14 and 28 dpa allowed lists of statistically significant, differentially expressed genes (DEG) between the transgenic and control lines to be compiled using an arbitrary cut-off of 1.5-fold (Supplementary Table S1). Only five out of the 9426



**Figure 1** High molecular weight (HMW) glutenin subunit profiles of the bread wheat lines used for transcriptome analysis. Lane 1, background, nontransformed L88-31 line; lane 2, transgenic B102-1-1 line; lane 3, conventionally bred L88-18 line; lane 4, background control nontransformed Cadenza (B1084-0-1) line; lane 5, 'clean' fragment-transgenic B1355-4-2(18) line; lane 6, 'whole plasmid'-transgenic B1118-8-4(6) line. The position of 1Ax1 protein (a) is shown, as well as other, nontransgenic endogenous HMW subunit proteins in the different wheat lines analysed: (b) Dx5, (c) Bx14/By15, (d) Bx17/By18 and (e) Dy10. The summary table describes the relevant gene composition of the different wheat lines under study.

Wheat Line	Endogenous HMW Subunits	Transgenes contained on: Reporter Plasmid	Marker Plasmid
<b>L88-31 control</b>	1Bx17+ 1By18	none	none
<b>L88-18 conv. bred</b>	1Ax1, 1Bx17+1By18	none	none
<b>B102-1-1</b>	1Bx17+ 1By18	1Ax1+ Amp <sup>R</sup>	bar, UidA, Amp <sup>R</sup>
<b>B1084-0-1 control</b>	1Bx14+1By15, 1Dx5+1Dy10	none	none
<b>B1355-4-2</b>	1Bx14+1By15, 1Dx5+1Dy10	1Ax1(fragment)	bar(fragment)
<b>B1118-8-4</b>	1Bx14+1By15, 1Dx5+1Dy10	1Ax1+ Amp <sup>R</sup>	bar, UidA, Amp <sup>R</sup>

**Table 1** Numbers of statistically significant and differentially expressed genes in the comparison of the three wheat lines B102-1-1, L88-31 and L88-18

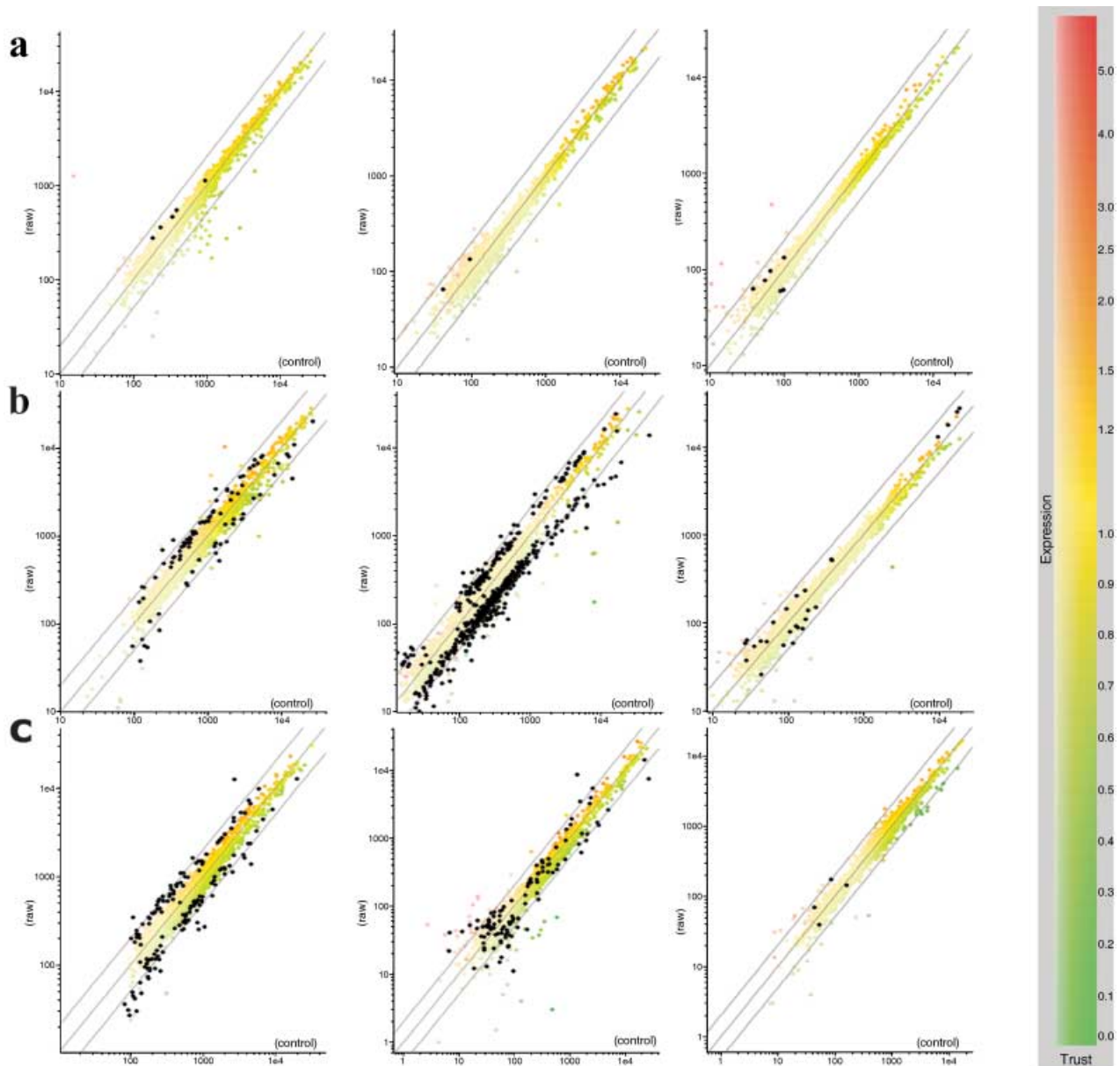
	14 day		28 day		8 day	
	No.	%	No.	%	No.	%
B102-1-1 vs. L88-31	5	0.05	2	0.02	6	0.06
L88-18 vs. L88-31	92	0.99	527	0.59	26	0.27
B102-1-1 vs. L88-18	154	1.63	118	1.25	4	0.04

The table shows the total number (also expressed as percentage of the arrayed 9426 unigene set) of statistically significant and differentially expressed genes at an arbitrary 1.5-fold cut-off for the different transcriptome comparisons: transgenic B102-1-1 line vs. nontransformed L88-31, conventionally bred L88-18 line vs. nontransformed L88-31 (sister line) and transgenic B102-1-1 line vs. conventionally bred L88-18. The transcriptome comparisons were performed at different developmental stages: 14 dpa (endosperm), 28 dpa (endosperm) and 8 dpv (leaf).

wheat ESTs represented in the microarray showed statistically significant differential expression between B102-1-1 vs. L88-31 at 14 dpa (Figure 2a, left), with none of the differences being greater than twofold. Four of these EST sequences corresponded to the same gene, identified as showing sequence

similarity to the *Triticum monococcum* ACT-1 gene (accession number AF326781), and differential expression of two of these sequences was confirmed by real-time RT-PCR (Supplementary Table S2a). The microarray data were also validated by quantitative RT-PCR of two types of genes: ones that did not show differential expression between the transgenic (B102-1-1) and control (L88-31) lines, and ones that showed low but statistically significant differential expression (see, for example, wheat chloroform/methanol-soluble protein CM2 in Supplementary Table S2a). Similar comparisons of the mRNA fractions from 28 dpa tissues (Figure 2a, middle) showed that only two genes were significantly ( $P < 0.05$ ) differentially expressed at a level above 1.5-fold. Quantitative real-time RT-PCR confirmed the differential expression of one of these (the *OSERS* gene) but showed no significant difference in the expression of the second gene (corresponding to an inorganic pyrophosphatase) (Supplementary Table S2a). Comparison of gene expression in the leaves of the two lines at 8 dpv identified only six statistically significant DEG (all at less than 1.5-fold) (Figure 2a, right), and in all cases the data were confirmed by real-time RT-PCR (Supplementary Table S2a).

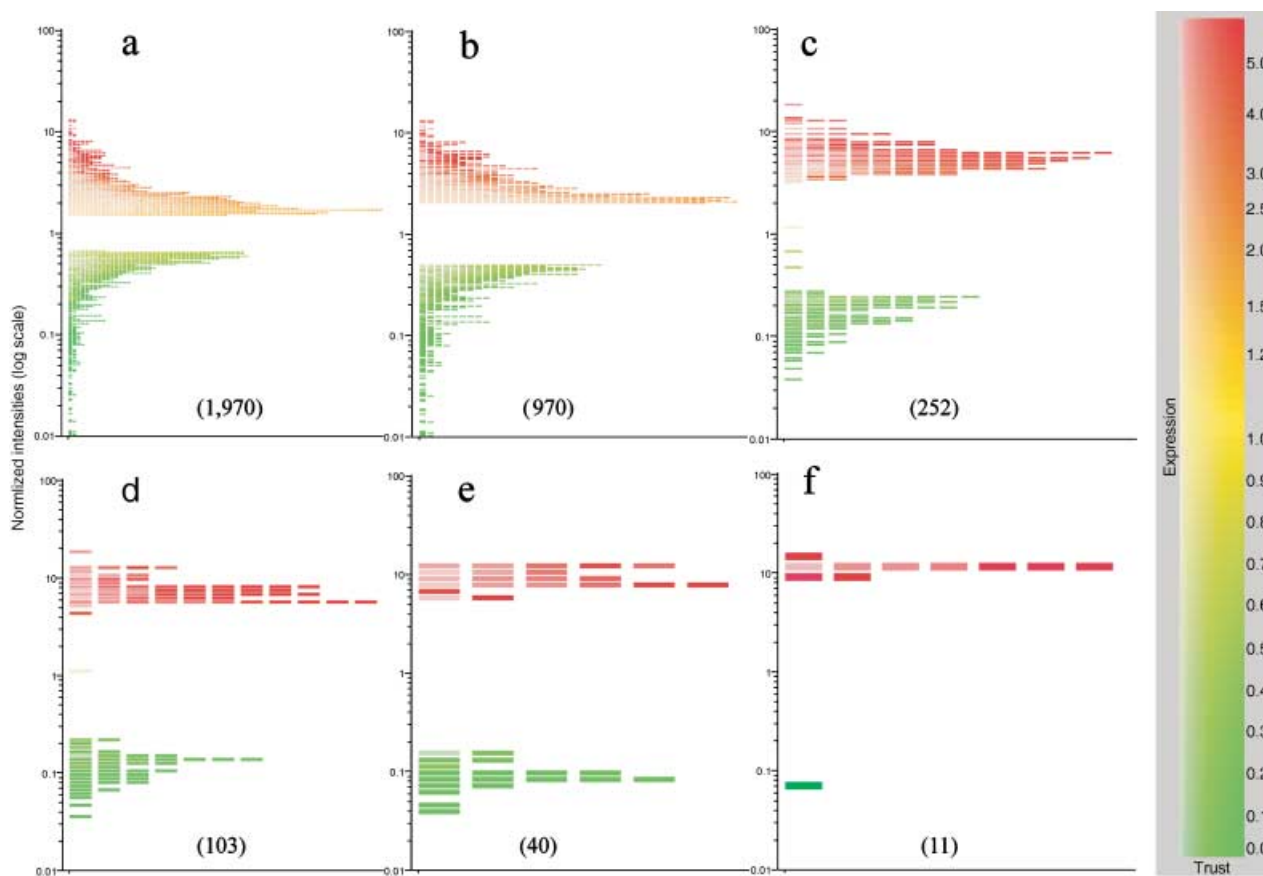
To place the differences observed between the expression profiles of the endosperms of the transgenic and nontransgenic



**Figure 2** Scatter plot representation of transcriptome comparisons of: (a) transgenic B102-1-1 line vs. control L88-31 line in endosperm at 14 dpa (left), 28 dpa (middle) or leaf at 8 dp (right); (b) conventionally bred L88-18 vs. L88-31 line in endosperm at 14 dpa (left), 28 dpa (middle), or leaf at 8 dp (right); (c) transgenic B102-1-1 line vs. conventionally bred L88-18 line in endosperm at 14 dpa (left), 28 dpa (middle), or leaf at 8 dp (right). Dots represent the normalized relative expression level of each arrayed gene for the transcriptome comparisons described. Dots in black represent statistically significant, differentially expressed genes (DEG) at an arbitrary cut off  $> 1.5$ . The inner line on each graph represents no change in expression. The offset dashed lines are set at a relative expression cut-off of twofold. In the adjacent coloured bar (rectangle on the far right of the figure), the vertical axis represents relative gene expression levels: reds indicate overexpression, yellows average expression, and greens under-expression. Values are expressed as  $n$ -fold changes. The horizontal axis of this bar represents the degree to which data can be trusted: dark or unsaturated colour represents low trust and bright or saturated colour represents high trust.

lines in a wider developmental context, the transcriptomes of endosperm at 14 dpa and leaf at 8 dp of the transgenic line B102-1-1 were compared to each other. A large number of genes (1970), representing 21% of the arrayed ESTs, were significantly ( $P < 0.05$ ) differentially regulated at expression levels higher than 1.5-fold. Furthermore, approximately 50%

(980) of these genes were up-regulated in the endosperm compared to the leaf, with the majority encoding proteins associated with endosperm development and storage product accumulation, including storage proteins and enzymes of starch biosynthesis. Similarly, many of the 990 genes, which were up-regulated in leaves, encoded proteins characteristic



**Figure 3** Histograms representing the transgenic line B102-1-1 transcriptome comparisons between endosperm at 14 dpa and leaf at 8 dpj using various arbitrary expression level cut-offs: (a) >1.5-fold, (b) >2-fold, (c) >4-fold, (d) >6-fold, (e) >8-fold, (f) 10-fold. Numbers in parentheses are the statistically significant DEG for the set cut-offs. In the adjacent coloured bar (rectangle on the far right of the figure), the vertical axis represents relative gene expression levels: reds indicate over-expression, yellows average expression, and greens under-expression. Values are expressed as n-fold. The horizontal axis of this bar represents the degree to which data can be trusted: dark or unsaturated colour represents low trust and bright or saturated colour represents high trust.

of photosynthetic tissues such as chlorophyll a/b binding protein, glutathione-S-transferase, PSI chloroplast precursor, carbonic anhydrase chloroplast precursor and thioredoxin. Histograms illustrating the expression data are given in Figure 3.

We also compared the transcriptomes of the nontransgenic lines L88-18 and L88-31 (Lawrence *et al.* 1988), which were selected from the same cross (Table 1 and Figure 2b). Comparison of the endosperms at 14 dpa showed that 92 genes (approximately 1% of the ESTs arrayed) were significantly ( $P < 0.05$ ) differentially regulated by 1.5-fold or greater (Figure 2b, left; compare to Figure 2a, left, for B102-1-1 vs. L88-31), of which, 14% (13 genes) showed differential expression of greater than threefold in L88-18 compared to L88-31 (Supplementary Table S3a). An even larger number of genes (527) showed statistically significant expression level differences greater than 1.5-fold when the endosperm transcriptomes of these lines were compared at 28 dpa (Figure 2b, middle; compare to Figure 2a, middle), with 16% (85) of

these being differentially expressed by more than twofold (Supplementary Table S3a). In contrast, comparison of the leaf transcriptomes of L88-18 and L88-31 at 8 dpj showed that only 26 genes differed significantly with expression levels at higher than 1.5-fold (Figure 2b, right; compare to Figure 1a, right for B102-1-1 vs. L88-31) (Supplementary Table S3a).

We also compared the transcriptomes of L88-18 (Lawrence *et al.* 1988) and B102-1-1 (Rooke *et al.*, 2003) (Figure 2c), which express the endogenous and transgenic forms of the 1Ax1 gene, respectively (Table 1 and Figure 1). These comparisons yielded 154 (1.6% of all genes analysed) and 118 (1.25%) genes, which showed statistically significant differential expression at levels between 1.5-fold and 4-fold in endosperms at 14 and 28 dpa, respectively (see Figure 2c, left and middle; Supplementary Table S3b). Of these genes, 25% (for endosperm at 14 dpa) and 36% (for endosperm at 28 dpa) corresponded to the same genes that were differentially expressed between the endosperm transcriptomes of L88-18

and L88-31, which is consistent with L88-31 being the parental line of B102-1-1. When the transcriptomes of 8 dpg leaves of B102-1-1 and L88-18 were compared only four statistically significant DEG could be identified that showed expression levels higher than the arbitrary cut-off of 1.5-fold (Figure 2c, right; Supplementary Table S3b). The lower numbers of differentially expressed genes in this comparison than in the comparisons of developing endosperms may relate to the fact that only the marker genes (*bar* and *UidA*) were expressed in the leaves of B102-1-1, not the HMW subunit transgenes.

Based on the results for the two comparisons of endosperm transcription profiles (L88-18 vs. L88-31 and B102-1-1 vs. L88-18), a total of 37 genes with statistically significant differential expression at levels higher than 1.5-fold were selected for validation of the microarray expression data by quantitative RT-PCR. Results obtained from this analysis confirmed the microarray data for 28 of the genes, the remaining nine showing no significant differences (Supplementary Table S2a).

#### Comparisons of lines transformed with clean fragments and whole plasmids

The wheat line B102-1-1 was transformed by biolistics (i.e. particle bombardment) using two whole plasmids: p1Ax1, which contains the *Amp<sup>R</sup>* gene controlled by the prokaryotic *npt1* promoter and the *Glu-A1* gene, and pAHC25, which has the *Amp<sup>R</sup>* gene and the *bar* and *UidA* genes controlled by the maize ubiquitin promoter. In addition to these genes, both plasmids also have the plasmid backbone sequences. The plasmid backbone, *UidA* and *Amp<sup>R</sup>* sequences are not required for the transformation, with the latter being of concern to regulatory authorities because of the potential for cross-kingdom transfer of antibiotic resistance to pests, pathogens and consumers. We therefore transformed the UK bread wheat cultivar Cadenza (B1084-0-1) with either the whole p1Ax1 and pAHC25 plasmids (line B1118-8-4) or with the *Glu-A1* and *bar* genes as 'clean fragments' which had been excised from their plasmids and purified (line B1355-4-2). Both lines were validated for transgene homozygosity, and B1355-4-2 was also validated for the absence of contaminating plasmid sequences. SDS-PAGE showed that the two transgenic lines expressed the 1Ax1 protein at approximately similar levels (Figure 1). These two lines were therefore used for transcriptional analyses with the control untransformed line Cadenza.

Transcriptome comparisons were made between the two transgenic lines and their common, nontransgenic background line, and between the transgenic lines themselves,

**Table 2** Numbers of statistically significant and differentially expressed genes for comparison of lines transformed with whole plasmids (B1118-8-4) or 'clean fragments' (B1355-4-2)

	14 day		28 day		8 day	
	No.	%	No.	%	No.	%
B13554 vs. Cadenza	6	0.06	9	0.1	1	0.01
B1118 vs. Cadenza	7	0.07	12	0.13	2	0.02
B13554 vs. B1118	26	0.28	4	0.04	3	0.03

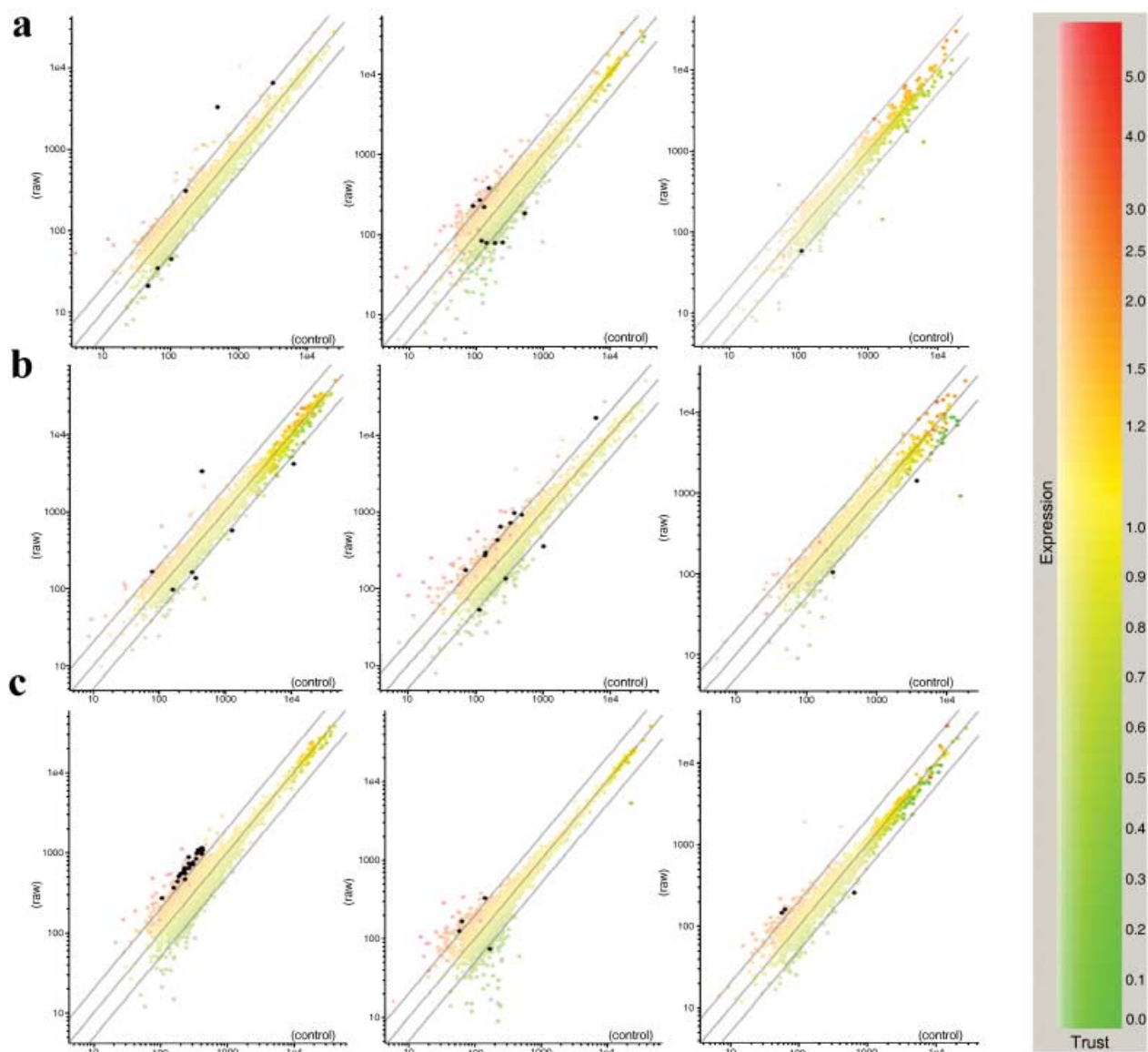
The table shows the total number (also expressed as percentage of the arrayed 9426 unigene set) of statistically significant and differentially expressed genes at an arbitrary 1.5-fold cut-off for the different transcriptome comparisons: transgenic B1355-4-2 ('clean fragments') or B1118-8-4 ('whole plasmids') line vs. nontransformed Cadenza and transgenic B1355-4-2 and B1118-8-4 compared between themselves. The transcriptome comparisons were performed at different developmental stages: 14 dpa (endosperm), 28 dpa (endosperm) and 8 dpg (leaf).

using RNA extracted from endosperm at 14 and 28 dpa, and from leaf tissue at 8 dpg (Figure 4).

The results of the transcriptome comparisons between B1355-4-2 (transformed with clean fragments), B1118-8-4 (whole plasmids) and Cadenza (B1084-0-1 control) are summarized in Figure 4 and Table 2 with further details being given in Supplementary Table S4 (a–c). In all cases the numbers of genes which showed statistically significant differential regulation were low, particularly in the leaf samples where they corresponded to only 0.01%–0.03% of the total genes arrayed. No clear trends were observed in the relative numbers of differentially regulated transcripts in the three comparisons. The differences in expression of selected transcripts between lines were confirmed by quantitative RT-PCR in most of the comparisons made (13 out of 17) (Supplementary Table S2b).

## Discussion

The modification of crops via genetic engineering is a subject of public concern. A question often asked is 'do genetically modified crops differ significantly from the nonmodified equivalent?' The term 'substantial equivalence' has been used in the area of food safety and biotechnology to describe the relationship between components produced from the same source using either novel or conventional methodologies. If resulting components are indistinguishable, they can be considered equivalent (FAO/WHO, 1996). Substantial equivalence in the context of the work reported here is used to describe the relationship between the phenotype of a genetically modified plant line and the originating variety.



**Figure 4** Scatter plot representation of transcriptome comparisons of: (a) transgenic B1355-4-2(18) line ('clean fragment') vs. control, nontransgenic Cadenza line (B1084-0-1) in endosperm at 14 dpa (left), 28 dpa (middle), or leaf at 8 dpg (right); (b) transgenic line B1118-8-4(6) ('whole plasmid') vs. control, nontransgenic Cadenza line (B1084-0-1) in endosperm at 14 dpa (left), 28 dpa (middle), or leaf at 8 dpg (right); (c) B1118-8-4(6) vs. B1355-4-2(18) transgenic lines in endosperm at 14 dpa (left), 28 dpa (middle), of leaf at 8 dpg (right). Dots, bar colour codes and expression level cut-offs are as in Figure 1.

Clearly, this type of comparison has emotive implications; however, the current view among the legislative authorities in Europe suggests that substantial equivalence can be established by demonstrating that the characteristics assessed for the genetically modified organism are equivalent to the same characteristics of the conventional comparator. In practice, such comparisons are difficult to make due to the levels and variation of phenotype found in natural material. Therefore, it has been agreed that to be substantially equivalent the genetically modified organism must be within the natural range of variation for those characteristics considered in the

comparator and be based upon an appropriate analysis of data (FAO/WHO, 1996). In the experiments described here, we have shown that several transgenic wheat lines carrying either specific gene fragments, or plasmids containing such fragments, show similar gene expression patterns to non-transgenic originator plant lines. We intentionally used a transgene with a well-characterized nonenzymatic function and highly specific developmental expression profile, in order to simplify analysis of substantial equivalence. A detailed statistical analysis of the results presented here demonstrated that differences in gene expression profiles between transgenic

and untransformed lines were extremely small (mostly less than 1.5-fold differences) when comparing the same tissues and developmental time points, and were in all cases below the arbitrary level used in many microarray studies to indicate differential expression of genes (usually set at twofold). The minor differences which were observed may have resulted from the transgenic and control lines not being genetically identical, due to a low level of residual heterozygosity in the material used for transformation.

Analysis of results obtained from experiments to compare the effects of using either whole plasmids or 'clean fragments' as the source of transgenes also demonstrated that both types of transgenic manipulation led to similar very small or insignificant changes in expression profiles, indicating that associated marker genes do not influence gene expression in the wheat lines and tissues studied here. Most strikingly, our results demonstrate far greater differences in expression profiles for comparisons of related lines of wheat produced by conventional breeding than for transgenic vs. untransformed lines. This implies that transgenesis (i.e. the presence of the transgene and associated introduced DNA) *per se* has a much smaller impact on global gene expression patterns than conventional alteration of the genome content through breeding. In addition, it implies that manipulation of phenotype via conventional crossing of lines results in derived material that could not be considered substantially equivalent in terms of the transcriptome sampling described here. Therefore, transgenesis can be considered to be a more precise mechanism for altering phenotype than can conventional breeding (although this of course depends on the function of the introduced transgene).

However, there are several important caveats to the interpretation of the data presented here. For each experiment, only one derived transgenic line was analysed for each construct tested. This was considered appropriate for this study as some of these lines have already been used in field-scale evaluations of performance (Darlington *et al.*, 2003; Shewry *et al.*, 2006). Furthermore, we only analysed the influence of a transgene with no known enzymatic or other metabolic activity (a glutenin storage protein). Analysis of transgenes that influence metabolomic or developmental pathways will therefore be required to provide information about the substantial equivalence at these levels of organization. Finally, we have not addressed the interactions between environmental variables, transgenes and substantial equivalence. It has been suggested that significant differences between transgenic and nontransgenic material might be observed under conditions of stress. Due to the complex nature of such experiments, we have not carried out a detailed analysis of

the lines under various environmental conditions. However, previous work by our group has shown that gene expression in wheat grain is significantly influenced by environmental conditions. For example, we have shown that developing wheat grain grown under field conditions with either organic or inorganic nitrogen had significantly different transcript profiles, with over 640 transcripts (> 6%) of array showing a greater than twofold difference (Lu *et al.*, 2005). Those effects are clearly much greater than the ones described here using wheat under controlled environmental conditions, with none of our transgenic/nontransgenic comparisons showing more than 1.6% of transcripts as having significantly different expression profiles. Further work in this area by El Ouakfaoui and Miki (2005) showed that the transcript profiles of transgenic and nontransgenic *Arabidopsis* material were indistinguishable in response to stress. Taken together, these results suggest that there are differences in gene expression levels between transgenic and nontransgenic crops grown under different environmental conditions, for instance, dry vs. wet, organic vs. inorganic, than between transgenic and nontransgenic crops grown under similar conditions.

This is, to our knowledge, only the second published comparison of the substantial equivalence of a GM and a non-GM crop at the transcriptome level and the first in which an endogenous wheat gene was used for transformation. Gregersen *et al.* (2005) reported similar studies on wheat expressing a fungal phytase gene using the same microarray as that used here. Only small differences were observed which the authors ascribed to minor asynchrony in the developmental stages of the seeds harvested for analysis.

Our results are also consistent with more detailed comparisons reported for field-grown grain of B102-1-1 and L88-31. Comparison of the metabolite profiles of grain grown on two sites over three seasons (1999–2001) failed to show significant differences in the compositions of the two lines (Baker *et al.*, 2006). Similarly, Shewry *et al.* (2006) showed no significant differences in the stability of HMW subunit gene expression or in the stability of grain weight, grain nitrogen content and dough strength in a similar series of samples (grown in 1998–2000).

It is clearly not possible to conclude that all transgenic crops are substantially equivalent to parental lines based on the limited transcriptome comparisons reported here. However, we have demonstrated that a single transgene has minimal effects on the transcriptome profiles of three lines when compared to the appropriate controls. Thus, these transgene/parental line combinations are substantially equivalent. This clearly establishes the principle that it is possible to produce transgenic crops which are substantially equivalent to the control nontransformed lines.

## Experimental procedures

### Plant material

Endosperms and leaves of three transgenic lines of hexaploid bread wheat (*Triticum aestivum*) were used for transcriptome comparison studies. The transgenic wheat lines B102-1-1 (Barro *et al.*, 1997; Rooke *et al.*, 2003) (in the L88-31 background) and B1118-8-4 (in the Cadenza background), were produced at Rothamsted Research by co-bombardment with the p1Ax1 plasmid (Halford *et al.*, 1992) containing the HMW glutenin subunit *1Ax1* (*Glu-1Ax*) gene under the control of its own endosperm-specific promoter, and a plasmid carrying the selectable *bar* gene (accession number X17220) and the marker gene *UidA* (accession number S69414) under the control of the maize ubiquitin promoter (Christensen and Quail, 1996). The transgenic line, B1355-4-2, also derived from Cadenza, was obtained by co-transformation using the 'clean' fragments corresponding to the HMW glutenin subunit *1Ax1* gene and the *bar* gene coding sequences. A conventionally bred, sister line to L88-31 line, L88-18 (Lawrence *et al.*, 1988), was also used for transcriptome comparison. The previously characterized transgenic line B102-1-1 (Barro *et al.*, 1997; Rooke *et al.*, 2003) and the transgenic Cadenza lines B1118-8-4(6) and B1355-4-2(18) were selected in the T2 generation for homogeneous, nonsegregating expression of the 1Ax1 protein by SDS-PAGE and for the presence of the *bar* gene by PCR. The expression of the HMW-GS 1Ax1 protein was determined for all wheat lines used (Figure 1) by SDS-PAGE of total protein extracted from grain using 10% (w/v) acrylamide gels and a Tris-borate buffer system as described in Shewry *et al.* (1995).

### Glasshouse experiments

Plants for all experiments were grown in containment glasshouse compartments under controlled environment conditions (18–20 °C/10–14 °C day/night; 16 h day/8 h night, 50%–70% humidity, 750 µE/s/m<sup>2</sup> irradiance) with automatic watering. Transgenic and control, nontransformed wheat lines selected for the different transcriptome comparison studies were grown in parallel under identical environmental conditions. Plants were grown in pots and arranged in a balanced row and column design, providing three biological replicates per treatment (wheat line and developmental stage) for microarray hybridizations. Two plants were grown per pot because two grain developmental stages were studied (14 and 28 dpa). Two tillers were kept per plant. A third plant was grown in selected pots for leaf tissue, sampling at 8 dpg. For the transcriptome comparisons between the transgenic

line B102-1-1 and the nontransformed, control line L88-31, the lines were grown as described but the experiment was performed at two different times of the year in the glasshouse. For the second experiment, the transgenic B102-1-1 and control, nontransgenic L88-31 lines were grown together with L88-18. Plants were visually monitored and tagged at the onset of anthesis, and caryopsis samples were collected at 14 and 28 dpa. For leaf samples, plants were harvested at 8 dpg. Seed endosperm was manually dissected under aseptic conditions, and both endosperm and expanded leaf tissues were used for total RNA extraction.

### RNA extraction

Total RNA was extracted from endosperms using the method of Chang *et al.* (1993). Total RNA from leaves was extracted following the method described by Cheng *et al.* (2001), except that the extraction buffer was replaced by SDS/phenol buffer. Also, much of the cell material complexed with SDS was removed from the homogenate before phenol/CHCl<sub>3</sub> extraction by the addition of 0.2 volume of 5 M KAc and centrifugation. Total RNAs were DNase treated (DNase Kit, Ambion, Cambridgeshire, UK) and the integrity of the RNA was checked in formaldehyde agarose gels and quantified using the Agilent 2100 Bioanalyser (RNA 6000 Nano LabChip Kit, Agilent Technologies, Palo Alto, CA).

### cDNA labelling

The cDNA probes used to hybridize microarray slides were generated from total RNA (from endosperm or leaf), reverse transcribed and secondary conjugated to amino-allyl fluor dyes Alexa 555 (green) and 647 (red) (Molecular Probes Inc., Carlsbad, CA, USA). The cDNA microarray slides were hybridized in reversed dye. Poly(A)<sup>+</sup> cDNA transcripts corresponding to mammal-specific mRNA sequences were used as spiking controls and obtained as recombinant DNAs in the plasmid vector pT7T3D-Pac from the maize Gene Discovery Project (<http://www.zmdb.iastate.edu/>). The list includes clones pHUM2 (GB: AA418251), pHUM4 (GB:AA464627), pHUM5 (GB:H28469), pHUM8 (GB:AA485668) and TRAF1 (GB:BC024145), from the UK Medical Research Council Gene Service, Babraham, UK.

After spiking with different control transcripts at various concentrations, total RNA samples were reverse transcribed using an oligo(dT)<sub>23</sub> anchor primer (Sigma-Genosis, Haverhill, UK) and Superscript III reverse transcriptase (Invitrogen, Paisley, UK) in the presence of 5-(3-aminoallyl)-2'-deoxyuridine 5' triphosphate (AA-dUTP) (Sigma) in order to produce amino-allyl-labelled first strand cDNAs (aa-dUTP-cDNA). The aa-dUTP-cDNAs were purified

using MiniElute spin columns (Qiagen, Crawley, UK) and equally divided for esterification to Alexa Fluor dyes. The ready-to-use succinimidyl esters of the Alexa dyes (Molecular Probes) were used to label the aa-dUTP-cDNAs following the manufacturer's instructions. After labelling, the fluor dye aa-dUTP-cDNAs were again purified using MiniElute spin columns (QIAGEN) and the elutant was used immediately for the array hybridizations.

### Microarray hybridizations

We used a wheat cDNA microarray of 19 846 spots containing 9246 unigene sequences (<http://www.cerealsdb.uk.net/index.htm>), including internal positive and negative controls, that was produced and described by Wilson *et al.* (2004). Duplicated unigene set arrays were also spotted on to Codelink-activated slides (Amersham Biosciences Ltd, Buckinghamshire, UK).

Arrays were hybridized in reverse dye labelling to fluor dye aa-dUTP-cDNA samples using Alexa Fluor dyes 555 and 647, as instructed by the slide manufacturer. After hybridization, slides were washed for 15 min in 2XSSC, 1% (w/v) SDS at Room Temperature (RT) followed by 8 min in 1XSSC, 0.2% (w/v) SDS at RT, and 5 min in 0.1XSSC, 0.2% (w/v) SDS at RT. The slides were then immediately dried by a brief centrifugation at 450 g.

Hybridizations were performed as single-pair-comparisons between the transgenic wheat line (B102-1-1 or B1118-8-4 or B1355-4-2) and its background, nontransgenic line (L88-31 or Cadenza) at two stages of endosperm development (14 and 28 dpa), and with leaf tissue at 8 dp. Hybridizations were also performed for transcriptome comparisons between the conventionally bred wheat sister line L88-18 and L88-31 or the transgenic line B102-1-1. Each of these comparisons was made using three biological replicates for the line, tissue and developmental stage selected. A further experiment was performed to compare the transcriptomic profiles of endosperm at 14 dpa and leaf at 8 dp, both from the transgenic wheat line B102-1-1.

The hybridized slides were scanned using the Axon Instruments GenePix 4000B dual laser microarray scanner, and the extracted raw data (mean ratios) were analysed using GenePix version 5 software.

### Microarray data analysis

Data for each comparison were imported from GenePix to GeneSpring (GeneSpring 6.2, Silicon Genetics, USA) for normalization using a locally weighted scatter plot smooth (LOWESS) method to ensure that no undesirable relationship in the  $\log_2$  dye ratio (differential expression) vs.  $\log_2$  dye prod-

uct (intensity) plots occurred. With reference to methods discussed by Kerr (2003), the normalized data were then analysed using GenStat (GenStat 7th Edition, GenStat Procedure Library, Release PL15, Lawes Agricultural Trust, Rothamsted, Harpenden, UK) through fitting a linear mixed model consisting of random and fixed terms (see, for example, Searle *et al.*, 1992), by the method of Residual Maximum Likelihood (REML). This took account of terms relating to the variance structure of the data (i.e. the random or design terms): experiment (two for the B102-1-1 vs. L88-31 comparison), biological replicate and technical replicate (dye swap); and included a fixed term for genes. Using 'experiment' as a fixed rather than a random model term allowed a test to be made as to whether, when pooling the data from the two separate experiments for B102-1-1 vs. L88-31, the experiment effect was due to an actual treatment effect (i.e. due to time of year) or to just design variation, the test concluding the latter (for details see Supplementary material supporting Microarray Data Analysis). Comparison of model deviance (McCulloch and Nelder, 1989) confirmed the best form for the random part of each model, the change in deviance being asymptotically distributed as chi-squared on the degrees of freedom associated with the model change. The inclusion of terms therefore reflects that they were significant ( $P < 0.05$ ) at explaining variability in the normalized data. The models used for each experimental comparison by wheat material/time point are given in Supplementary Table S5. The ratio of each gene effect to its standard error, giving a  $t$ -statistic on the model residual degrees of freedom, enabled the significance of the differential expression from 0 (on the  $\log_2$  scale) to be assessed. Genes with significant differential expression ( $P < 0.05$ ) were filtered on expression (> 1.5-fold, > 2-fold > 3-fold, etc.) and then on the number of replicates present (nonmissing observations) for each gene. Genes represented by two or more replicates were retained and considered for real-time RT-PCR (reverse transcriptase-polymerase chain reaction), these having most information from the microarray as regards differential expression. The statistical methodology avoided the problem of multiple testing, i.e. by testing genes with respect to the overall residual variation in the data given a model, having accounted for significant design terms. Details of statistical models used for analysis of transformation methods using Cadenza material are included in Supplementary material supporting Microarray Data Analysis.

### Relative quantitative real-time RT-PCR

A list of statistically significant, differentially expressed genes (DEG) from the different transcriptome comparisons

was compiled for microarray data validation by real-time RT-PCR.

Specific pairs of primers for SYBR-green detection and quantification of selected DEG for microarray validation (EST clone sequences searched in <http://www.cerealsdb.uk.net/index.htm>) were designed using Primer Express® software (ABI® PRISM, Applied Biosystems, Foster City, CA, USA) following the TaqMan® Probe and Primer Design guidance. Dual-fluorogenic probes and primers (Prologo, <http://www.prologo.com>) were used for the detection and relative quantification of the specific *HMW-Glutenin Subunit 1Ax1* transgene (forward primer: 5'-AACCAACACAAGGGCAACAAAG-3'; reverse primer: 5'-GCTGCGAAGAAGTTGGATAGTATC-3'; probe: 5'-[Hex]-AGCAACCAGGACAAGGTCAGCAACTAAGA-[BHQ1]-3'), and for *bar* (forward primer: 5'-ACCACTACATC-GAGACAAGCACG-3'; reverse primer: 5'-ACGAGGTCGTC-CGTCCACTC-3'; probe: 5'-[Hex]CGTACCGAGCCGCAGGAA-CCGCA[BHQ1]-3') and *UidA* (forward primer: 5'-ATGCTCT-ACACCACGCCGAA-3'; reverse primer: 5'-AGTTCAACGCT-GACATCACCAT-3'; probe: 5'-[Hex]CCACCACCTGCCAGT-CAACAGACGC[BHQ1]-3') marker genes from different treatments (line/tissue/developmental stage) studied. When real-time RT-PCR assays were performed, an additional control PCR reaction was run in parallel using primers specific for the endogenous *actin* gene to normalize the quantified expression levels of the target DEG [the *actin* corresponds to clone ID H01\_p335\_plate\_6 in <http://www.cerealsdb.uk.net/index.htm> (accession number AAA33433); forward primer: 5'-AGGCATCCTGACGCTCAAGTA-3'; reverse primer: 5'-GCTCGTTGTAGAAGGTGTGGTG-3'; probe: 5'-[6Fam]CTC-CATGTCGTCCAGTTGCCAC[C[BHQ1]-3']. The expression of this *actin* gene was observed not to be modulated in the different microarray hybridization experiments performed.

Polymerase chain reactions were performed in optical 96-well plates with an ABI PRISM 7500 Sequence Detection System (Applied Biosystems), using Platinum Quantitative PCR SuperMix-UDG (Invitrogen) when TaqMan probe was used or Platinum SYBR® Green qPCR PCR SuperMix-UDG (Invitrogen). Total RNA was reverse transcribed using the Reverse Transcriptase (RT)-Enzyme mix (Invitrogen) containing SuperScript™ III RT and RNaseOUT. The 2XRT-Reaction Mix (Invitrogen) contained 2.5 µM oligo (dT)<sub>20</sub>, 2.5 ng/µL random hexamers and 10 mM MgCl<sub>2</sub> and dNTPs. The PCRs were performed with 100 ng cDNA, 12.5 µL of 2X Platinum® qPCR Super Mix-UDG [with or without SYBR Green, containing 60 U/mL Platinum *Taq* DNA polymerase, 40 mM Tris-HCl (pH 8.4), 100 mM KCl, 6 mM MgCl<sub>2</sub>, 400 µM dGTP, 400 µM dATP, 400 µM dCTP, 400 µM dUTP, 40 U/mL UDG and stabilizers] (Invitrogen), 0.5 µL ROX Reference Dye 0.5 µL (Invitro-

gen) and 200 ng of each specific primer in a final volume of 25 µL. A master mix with sufficient cDNA and reaction components was prepared prior to dispensing into individual wells to ensure that each reaction (i.e. two technical replicates per each of the three biological replicates per sample tested) contained an equal amount of cDNA to reduce pipetting errors. The following standard thermal profile was used for all PCRs: one cycle at 50 °C for 2 min; one cycle at 95 °C for 2 min; 40 cycles of 95 °C for 15 s and 60 °C for 1 min. A dissociation curve analysis was done for each pair of specific primers to detect if nonspecific amplification occurred. To generate a baseline-subtracted plot of the logarithmic increase in fluorescence signal ( $\delta R_n$ ) vs. cycle number, the baseline data were collected for most of the amplifications between cycle 3 and 15. Threshold cycle (Ct) values were collected from each sample. In order to compare the Ct values from different cDNA samples, the Ct of all the validated genes were normalized to the Ct value of the housekeeping gene *actin* (clone ID H01\_p335\_plate\_6; <http://www.cerealsdb.uk.net/index.htm>), which showed no differential expression between lines, tissues and developmental stages under study. PCR efficiency was estimated using the equations proposed by Ramakers *et al.* (2003). The relative expression of the selected genes for validation was calculated using the comparative method available in the ABI PRISM® SDS version 1.2.1 software (ABI PRISM, 2001) when the efficiency of amplification of the target and reference genes was similar. If these estimated efficiencies were different, the set of equations proposed by Pfaffl (2001) were used.

## Acknowledgements

This work was funded by the Biotechnology and Biological Sciences Research Council, UK (BBSRC) under the Gene Flow Agri-Food Initiative (ref. GM114152). The transgenic lines in cultivar Cadenza were produced as part of BBSRC LINK Project RMQ07453. We thank the following colleagues at Rothamsted Research: Professor Robin Thompson for help with statistical experimental design, Dr Rowan Mitchell for discussions on data analysis and Ms Melloney McClusky for assistance with plant growth and harvesting. We also thank Dr Greg Lawrence (CSIRO, Canberra, Australia) for providing the L88 series of lines and information on their production and Dr Bill Angus (Nickerson UK Ltd, Bury St. Edmunds, UK) for advice on conventional plant breeding practices.

## References

ABI PRISM (2001) *7700 Sequence Detection System: Relative Quantification*, Vol. 2. Foster City, CA: Applied Biosystems.

- Baker, J.M., Hawkins, N.D., Ward, J.L., Lovegrove, A., Napier, J.A., Shewry, P.R. and Beale, M.H. (2006) A metabolomic study of substantial equivalence of field-grown genetically modified wheat. *Plant Biotechnol. J.* **4**, 381–392.
- Barro, F., Rooke, L., Bekes, F., Gras, P., Tatham, A.S., Fido, R.J., Lazzeri, P., Shewry, P.R. and Barcelo, P. (1997) Transformation of wheat with HMW subunit genes results in improved functional properties. *Nat. Biotechnol.* **15**, 1295–1299.
- Chang, S., Puryear, J. and Cairney, J.A. (1993) Simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* **11**, 113–116.
- Cheng, G.P., Wilson, I.D., Kim, S.H. and Grierson, D. (2001) Inhibiting expression of a tomato ripening-associated membrane protein increases organic acids and reduces sugar levels of fruit. *Planta*, **212**, 799–807.
- Christensen, A.H. and Quail, P.H. (1996) Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Res.* **5**, 213–218.
- Darlington, H., Fido, R., Tatham, A.S., Jones, H., Salmon, S.E. and Shewry, P.R. (2003) Milling and baking properties of field grown wheat expressing HMW subunit transgenes. *J. Cereal Sci.* **38**, 301–306.
- El Ouakfaoui, S. and Miki, B. (2005) The stability of the *Arabidopsis* transcriptome in transgenic plants expressing the marker genes *ntp1l* and *uidA*. *Plant J.* **41**, 791–800.
- Evans, L.T. (1993) *Crop Evolution, Adaptation and Yield*. Cambridge: Cambridge University Press.
- FAO/WHO (1996) *Biotechnology and Food Safety: Report of a Joint FAO/WHO Consultation*. Rome: FAO.
- Gregersen, P.L., Brinch-Pedersen, H. and Holm, P.B. (2005) A microarray-based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. *Transgenic Res.* **14**, 887–905.
- Halford, N.G., Field, J.M., Blair, H., Urwin, P., Moore, K., Robert, L., Thompson, R., Flavell, R.B., Tatham, A.S. and Shewry, P.R. (1992) Analysis of HMW glutenin subunits encoded by chromosome 1A of bread wheat (*Triticum aestivum* L.) indicates quantitative effects on grain quality. *Theor. Appl. Genet.* **83**, 373–378.
- Kerr, M.K. (2003) Linear models for microarray data analysis: Hidden similarities and differences. *J. Comput. Biol.* **10**, 891–901.
- Lawrence, G.J., Macritchie, F. and Wrigley, C.W. (1988) Dough and baking quality of wheat lines in glutenin subunits controlled by *Glu-A1*, *Glu-B1* and *Glu-D1* loci. *J. Cereal Sci.* **7**, 109–112.
- Lu, C., Hawkesford, M.J., Barraclough, P.B., Poulton, P.R., Wilson, I.D., Barker, G.L. and Edwards, K.J. (2005) Markedly different gene expression in wheat grown with organic or inorganic fertilizer. *Proc. Biol. Sci.* **272**, 1901–1908.
- McCulloch, P. and Nelder, J.A. (1989) *Generalized Linear Models*. London: Chapman & Hall.
- Payne, P.I. (1987) Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Annu. Rev. Plant Physiol.* **38**, 141–153.
- Pfaffl, M.W. (2001) A new mathematical method for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, 2003–2007.
- Ramakers, C., Ruijter, J.M., Le Kanne Deprez and Moorman, A.F.M. (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction PCR (data). *Neurosci. Lett.* **339**, 62–69.
- Rooke, L., Steele, S.H., Barcelo, P., Shewry, P.R. and Lazzeri, P. (2003) Transgene inheritance, segregation and expression in bread wheat. *Euphytica*, **129**, 301–309.
- Searle, S.R., Casella, G. and McCulloch, C.E. (1992) *Variance Components*. London: Wiley.
- Shewry, P.R., Halford, N.G., Tatham, A.S., Popineau, Y., Lafiandra, D. and Belton, P.S. (2003) The high molecular weight subunits of wheat glutenin and their role in determining wheat processing properties. *Adv. Food Nutr. Res.* **45**, 221–302.
- Shewry, P.R., Powers, S., Field, J.M., Fido, R.J., Jones, H.D., Arnold, G.M., West, J., Lazzeri, P.A., Barcelo, P., Barro, F., Tatham, A.S., Bekes, F., Butow, B. and Darlington, H. (2006) Comparative field performance over three years and two sites of transgenic wheat lines expressing HMW subunit transgenes. *Theor. Appl. Genet.* in press.
- Shewry, P.R., Tatham, A.S. and Fido, R.J. (1995) *Plant Gene Transfer and Expression Protocols: Separation of Plant Proteins by Electrophoresis*, Vol. 49. Totowa, NJ: Humana Press.
- Wilson, I.D., Barker, G.L.A., Beswick, R.W., Shepherd, S.K., Lu, C., Coghill, J.A., Edwards, D., Owen, P., Lyons, R., Parker, J.S., Lenton, J.R., Holdsworth, M.J., Shewry, P.R. and Edwards, K.J. (2004) A transcriptomics resource for wheat functional genomics. *Plant Biotechnol. J.* **2**, 495–506.
- Wilson, I.D., Barker, G.L.A., Lu, C., Coghill, J.A., Beswick, R.W., Lenton, J.R. and Edwards, K.J. (2005) Alteration of the embryo transcriptome of hexaploid winter wheat (*Triticum aestivum* cv. Mercia) during maturation and germination. *Func. Integr. Genomics*, **5**, 144–154.

## Supplementary material

The following supplementary material is available online from <http://www.blackwell-synergy.com>

**Appendix S1** Details for analysis of comparison of the two experiments carried out for B102-1-1 and L88-31.

**Table S1** Transcriptome data analysis using mixed modeling method for transcriptome comparisons.

**Table S2** Microarray data validation by real-time RT-PCR.

**Table S3** Transcriptome data analysis using mixed modeling method for transcriptome comparisons.

**Table S4** Transcriptome data analysis using mixed modeling method for transcriptome comparisons.

**Table S5** The terms in the models found best for describing the log<sub>2</sub> normalised microarray data for each comparison being made.