Plant Biotechnology Journal (2004) **2**, pp. 495–506 doi: 10.1111/j.1467-7652.2004.00096.x

A transcriptomics resource for wheat functional genomics

Ian D. Wilson[1], Gary L. A. Barker[1], Richard W. Beswick[1,†], Sophie K. Shepherd[1,†], Chungui Lu[1,2], Jane A. Coghill[1], David Edwards[1,†], Philippa Owen[1], Rebecca Lyons[2], Jill S. Parker[1], John R. Lenton[1], Michael J. Holdsworth[1,2,†], Peter R. Shewry[2] and Keith J. Edwards[1,*]

[1]*University of Bristol, Department of Biological Sciences, Woodland Road, Bristol BS8 1UG, UK*
[2]*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

Received 10 April 2004;
revised 2 June 2004;
accepted 2 June 2004.
**Correspondence* (fax +44 (0)117 925 7374; e-mail k.j.edwards@bristol.ac.uk)
†*Present addresses*: Richard W. Beswick is currently at the Department of Haematology, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK; Sophie K. Shepherd is at the University of Reading, Department of Agriculture, Whiteknights, Reading RG6 6AH, UK; David Edwards is at Primary Industries Research Victoria, Department of Primary Industries, La Trobe University, Bundoora, Victoria 3086, Australia; and Michael J. Holdsworth is at the University of Nottingham, School of Biosciences, Division of Agricultural and Environmental Sciences, Sutton Bonington Campus, Loughborough, Leicestershire LE12 5RD, UK.
Keywords: cDNA, EST, microarray, transcriptome, Triticum aestivum.

Summary

Grain development, germination and plant development under abiotic stresses are areas of biology that are of considerable interest to the cereal community. Within the Investigating Gene Function programme we have produced the resources required to investigate alterations in the transcriptome of hexaploid wheat during these developmental processes. We have single pass sequenced the cDNAs of between 700 and 1300 randomly picked clones from each of 35 cDNA libraries representing highly specific stages of grain and plant development. Annotated sequencing results have been stored in a publicly accessible, online database at http://www.cerealsdb.uk.net. Each of the tissue stages used has also been photographed in detail, resulting in a collection of high-quality micrograph images detailing wheat grain development. These images have been collated and annotated in order to produce a web site focused on wheat development (http://www.wheatbp.net/). We have also produced high-density microarrays of a publicly available wheat unigene set based on the 35 cDNA libraries and have completed a number of microarray experiments which validate their quality.

Introduction

Wheat, with a global production approximating to 600 million tonnes per annum, together with rice and maize, dominates world agriculture (Aquino *et al*., 1999). Although the highest wheat yields have been achieved in Europe, limitations in its intrinsic processing quality, with strong year-to-year environmental effects, have necessitated the continued development of new cultivars more able to perform consistently under the prevailing conditions. There is, therefore, a real necessity to provide additional tools whereby the development of improved wheats can be achieved, in particular to increase the plant breeder's success in selecting for versatile, superior varieties suitable for factory-scale premium bread making. Given this, it is clear that a priority goal for academics and breeders is to better understand the interactions that occur between the wheat genome and environmental factors and to determine how these can be manipulated by agricultural practices, marker-directed breeding and genetic manipulation to produce a more consistent high-quality crop without loss of yield.

The advent of new molecular genetic technologies and the dramatic increase in plant gene sequence data have provided opportunities to underpin wheat breeding programmes in order to improve yield, grain quality and disease resistance (Langridge *et al*., 2001; Appels *et al*., 2003). Many of these technologies have been designed to facilitate the detection and understanding of the alterations in gene expression that accompany differential development or that result from the perception of changes to the environment. Within these contexts the goal has been to facilitate improvement either by providing genetic markers for breeding programmes or by enabling the application of genetic engineering technologies. To achieve a useful level of understanding about the function of a particular gene it is essential not only to study its temporal and spatial expression, but also those of other genes that may

be similarly coregulated and have effects on the phenotype. Thus, the ability to observe global and diagnostic changes in the mRNA population of an organism under examination has become paramount. Within the field of plant functional genomics the use of microarrays to profile the transcriptome (Ruan *et al*., 1998; Lockhart and Winzeler, 2000) has become a popular approach (Holtorf *et al*., 2002; Wilson *et al*., 2003). As the number of sequenced genes from different species continues to increase the use of microarrays to investigate aspects of plant biology has also expanded. Of the many recent studies, plant microarray-based investigations have profiled the transcriptome changes of *Arabidopsis thaliana* in different organs (Ruan *et al*., 1998), during seed development and germination (Girke *et al*., 2000; Ogawa *et al*., 2003), in response to changes in nitrate availability (Wang *et al*., 2000, 2003) and to pathogen attack (Narusaka *et al*., 2003). Studies of other plant species have profiled the transcriptomes of ripening strawberries (Aharoni *et al*., 2000), salt-stressed rice (Kawasaki *et al*., 2001), developing maize embryos (Lee *et al*., 2002), maize during cold acclimation (Kollipara *et al*., 2002) and following drought stress (Yu and Setter, 2003) and barley subjected to iron deficiency (Negishi *et al*., 2002).

The collation of large numbers of independently produced *A. thaliana* microarray data sets into publicly accessible, online databases, such as those held at NASC (Nottingham Arabidopsis Stock Centre; http://nasc.nott.ac.uk/), SIGnAL (Salk Institute Genomic Analysis Laboratory; http://signal.salk.edu/) or TAIR (The Arabidopsis Information Resource; http://www.arabidopsis.org/home.html), has done much to demonstrate how this approach can rapidly expand our knowledge of the biology of this and related species. However, although microarray resources are available for maize (http://www.maizegdb.org/) and there is also a USDA (US Department of Agriculture)-funded repository (http://barleypop.vrac.iastate.edu/BarleyBase/) for microarray data from the barley Affymetrix (http://www.affymetrix.com) Genechip, there is as yet no similar publicly available information for wheat. The transcriptome resources described here were initiated to improve publicly available academic resources for the wheat functional genomics community.

Results

Isolation of plant material, RNA extractions and library synthesis

Microscopic examination of individual Mercia wheat ears showed that, under the controlled environment growth conditions described (see 'Full experimental procedures' in supplementary data), anthesis (first shedding of pollen) occurred 3–5 min before the visible emergence of anthers from the ears. Tissues were sampled at defined times post-anthesis from the developing grains of the two lower florets of the first spikelets containing flowers visibly undergoing anther emergence. The tissues isolated from developing grains and other stages of wheat vegetative and reproductive development are listed in Table 1. Tissues from each time point were fixed and stained for microscopy as described and were used to produce a micrographic record of the developmental stages used to generate the transcriptomics resources described below. A pictorial, web-based, micrograph database describing these tissues and their developmental biology can be viewed at http://www.wheatbp.net/ and a selection of micrographs illustrating aspects of grain development is shown in Figure 1.

Total RNA and, subsequently, polyadenylated (polyA+) mRNA were successfully extracted from all the tissues listed in Table 1. Although, in each case, the quality of the total RNA extracted appeared to be equivalent, as determined by the appearance of discrete rRNA bands after agarose gel electrophoresis (data not shown), the yields on a per fresh weight basis varied considerably depending on the tissue in question. In general, the yields of RNA from the different tissues of early developing grain were considerably higher than those from the tissues of grain undergoing later development (see Table 1). In particular, there was a marked decline in the extractable RNA content of the grain endosperm 21 days post-anthesis (dpa) to the extent that, at 30 and 40 dpa, it proved impossible to obtain useful quantities from the tissue. There was also a general decline in the extractable RNA content of the developing embryo with time. Similarly, the extractable RNA content of the surrounding maternal tissue also declined after 14 dpa as the tissue senesced.

Libraries resulting from a single ligation of an aliquot of cDNA were produced representing each of the tissues listed (see Table 1). Depending on the library in question, the percentage of non-recombinants ranged from 3% to 13%, whilst the average library insert size ranged between 0.9 and 1.5 kilobase pairs (see Table 1).

Expressed sequence tag (EST) sequencing, data handling and analysis and unigene set production

Colonies picked from each library containing plasmid DNAs were used in single-pass sequencing reactions. In total 22 469 'good-quality' sequences, attaining a PHRED quality score of at least 20 over a minimum length of 100 base pairs, resulted from approximately 38 000 sequencing reactions in our laboratory. The overall sequencing success rate using these parameters was approximately 60%. To these sequences we

Table 1 Wheat cDNA libraries made or used within the Investigating Gene Function (IGF) programme with quality control checks

Tissue type	≈ RNA yield (μg/g fwt)	Library name	≈ cfu per 50 ng of ligated cDNA	≈ % library non-recombinants and ≈ average insert length (kbp)	No. of good sequences generated
Grain development/germination series					
Carpel 1 dpa	3000	A1:1	45 000	7%/1.5 kbp	786
Carpel 2 dpa	3000	A2:2	55 000	5%/1.5 kbp	841
Embryo enriched 6 dpa	2500	D1:5	26 000	8%/1 kbp	932
Embryo enriched 8 dpa	2500	D2:6	20 000	6%/1.2 kbp	440
Embryo enriched 10 dpa	1800	D3:7	24 000	9%/1 kbp	480
Endosperm 8 dpa	1500	E2:9	60 000	3%/1.5 kbp	749
Endosperm 10 dpa	1000	E3:10	120 000	7%/1.3 kbp	753
Maternal 6 dpa	10	E4:11	18 000	11%/0.8 kbp	669
Maternal 8 dpa	5	E5:12	16 000	12%/0.8 kbp	666
Endosperm 14 dpa	300	H1:16	100 000	4%/1.5 kbp	849
Embryo 14 dpa	1800	J1:22	66 000	7%/1 kbp	927
Embryo 21 dpa	1400	J2:23	70 000	7%/1 kbp	958
Embryo 28 dpa	1200	J3:24	35 000	6%/1 kbp	514
Embryo 30 dpa	1000	L1:25	42 000	8%/1 kbp	579
Embryo 40 dpa	1000	L2:26	28 000	7%/1 kbp	748
Embryo 1 dpg	1500	N1:29	60 000	10%/0.9 kbp	948
Embryo 2 dpg	1500	N2:30	50 000	8%/1 kbp	1 012
Whole grain minus embryo 2 dpg	5	O2:32	20 000	13%/0.7 kbp	747
Whole seedling (shoot/root) stress series					
Cold-acclimated	200	P1:33	36 000	10%/1 kbp	907
Drought-stressed	100	P2:34	22 000	6%/1 kbp	1 265
Salt-stressed	250	P3:35	43 000	8%/1.2 kbp	925
Waterlogged roots	30	P4:36	30 000	10%/1 kbp	715
Normal roots	50	P5:37	33 000	7%/1 kbp	448
Normal whole seedling	250	P6:38	45 000	10%/1 kbp	945
NO_3-starved whole seedling	50	P7:39	32 000	11%/0.9 kbp	662
NO_3-starved roots	20	P8:40	29 000	7%/1 kbp	420
Externally supplied libraries/sequences					
*Ovules 2 dpf	–	Q1:41	–	–	701
*Egg cell	–	Q2:42	–	–	654
*Unfertilized ovules	–	Q3:43	–	–	767
*2 Cell zygote	–	Q4:44	–	–	462
†Heat-stressed leaves	–	T05	–	–	901
†Heat-stressed leaves‡	–	T06	–	–	1 019
†Heat-stressed leaves‡	–	T07	–	–	1 002
†Drought-stressed leaves	–	T08	–	–	67
†Drought-stressed leaves‡	–	T09	–	–	924
Total					26 382

cfu, colony-forming units; dpa, days post-anthesis; dpg, days post-germination; dpf, days post-fertilization; fwt, fresh weight; kbp, kilobase pair; –, no data available.
*Libraries supplied by Dr S. Spruck, University of Hamburg, Germany.
†Sequences supplied by Dr N. Klueva, Texas Technical University, USA.
‡Subtracted library.

added 3913 sequences supplied by Dr N. Klueva of the Texas Technical University, USA.

Automated BLAST searches were performed on the resulting sequence data using the NCBI's network BLAST client (National Center for Biotechnology Information; blastcl3). Of the 26 382 sequences entered into similarity BLASTX searches 15 022 (57%) encoded proteins which were similar (expect = $1e^{-5}$) to others encoded by sequences with defined annotation whereby a reasonable attempt at assigning a biological role or function could be made. Of the remaining 11 360 sequences, 2087 (8%) encoded proteins similar to others in databases or to those defined simply as being expressed, predicted, unnamed, hypothetical or of unknown function. In the BLASTX searches the remaining 9273 (35%) sequences failed to produce translations

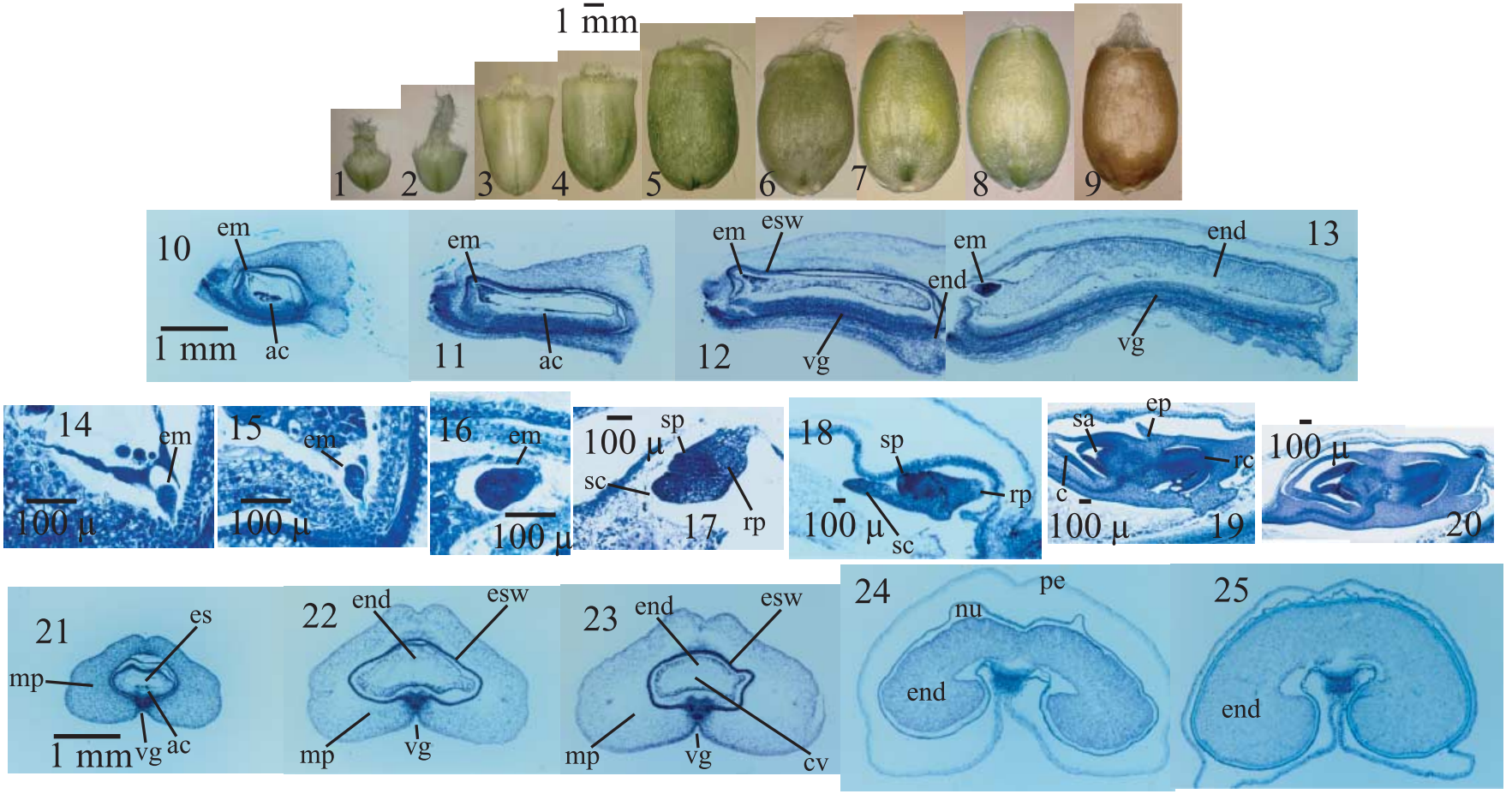

Figure 1 Whole grain and micrograph images of developing wheat grains. Dorsal views of the whole grain at: 1, 1 day post-anthesis (dpa); 2, 3 dpa; 3, 6 dpa; 4, 8 dpa; 5, 10 dpa; 6, 14 dpa; 7, 21 dpa; 8, 28 dpa; 9, 40 dpa. Longitudinal micrograph sections of the whole grain at: 10, 2 dpa; 11, 4 dpa; 12, 6 dpa; 13, 11 dpa. Micrographs showing development of the embryo at: 14, 2 dpa; 15, 5 dpa; 16, 7 dpa; 17, 11 dpa; 18, 15 dpa; 19, 21 dpa; 20, 28 dpa. Cross-section micrographs of the whole grain showing endosperm development at: 21, 2 dpa; 22, 5 dpa; 23, 7 dpa; 24, 15 dpa; 25, 21 dpa. ac, antipodal cells; c, coleoptile; cv, central vacuole; em, embryo; end, endosperm; ep, epiblast; es, embryo sac; esw, embryo sac wall; mp, maternal pericarp; nu, nucellus; pe, pericarp; rc, root cap; rp, root pole; sa, shoot apex; sc, scutellum; sp, shoot pole; vg, ventral groove.

with any significant similarities to protein sequences held within the NCBI non-redundant data set. However, of these 9273 sequences 2612 (10%) had significant (expect ≤ $1e^{-5}$) BlastN similarities to other sequences encoding proteins of known function. Thus, no sequence similarity or function could be assigned to 6661 (25%) of the sequences produced. The 15 022 sequences encoding proteins of known function, as determined by BlastX similarity searches, could be further subdivided into smaller groups. Of these, approximately 9.1% encoded proteins involved in transcription; 1.4% encoded proteins involved in DNA synthesis and cellular biogenesis, development, growth and division; 8.5% encoded proteins involved in cell rescue, defence, death and ageing; 5% encoded proteins involved with cellular transport and transport mechanisms; 7.5% encoded proteins involved in cellular communication and signal transduction; 10% encoded proteins involved in metabolism and energy production; and 15.5% encoded proteins involved in protein synthesis and targeting. Digital differential display and virtual Northern tools were developed to mine the cerealsdb data set for potentially differentially expressed sequences. An example of virtual Northern data showing the seedling stress-specific expression of mRNAs encoding a dehydrin (cerealsdb clone ID: E12_p133_plate_14; GenBank acc. no. AL821594), the late endosperm developmental expression of a γ-gliadin encoding mRNA (cerealsdb clone ID: H08_h116_plate_10; GenBank acc. no. AL874434) and the specific expression of an mRNA encoding a protein of unknown function (cerealsdb clone ID: C11_q242_plate_2; GenBank acc. no. AL831184) in the unfertilized egg cells is shown in Figure 2.

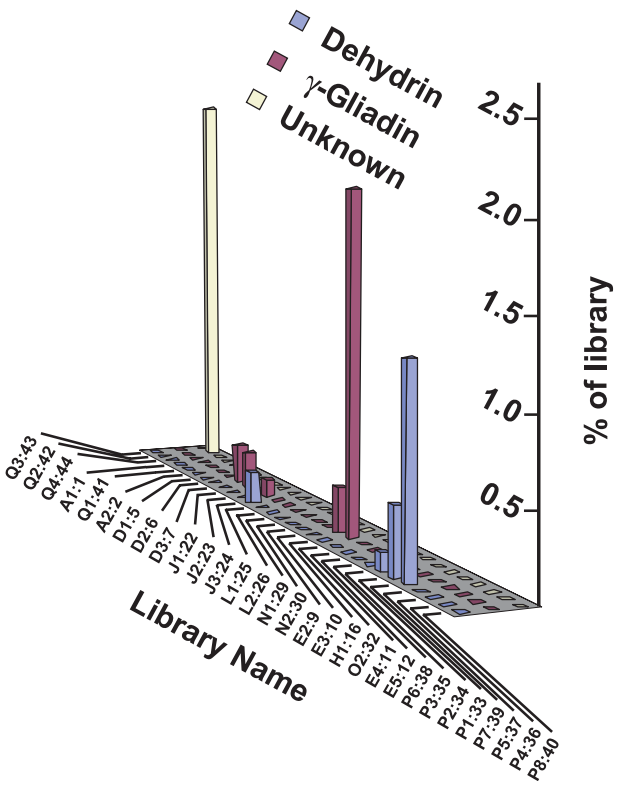

Figure 2 Virtual Northern showing the percentage of each library represented by mRNAs encoding a dehydrin (cerealsdb clone ID: E12_p133_plate_14; GenBank acc. no. AL821594), a γ-gliadin (cerealsdb clone ID: H08_h116_plate_10; GenBank acc. no. AL814434) and a protein of unknown function (cerealsdb clone ID: C11_q242_plate_2; GenBank acc. no. AL831184).

A unigene set was selected from sequences derived solely from within the cDNA libraries produced. In this instance 19 829 EST 'good-quality' sequences in FASTA format were aligned using cap3 with a similarity threshold of 85%. The resulting 9155 unigene sequence set was derived from 6353 singletons and 2802 contigs. In the case of contigs, the individual clone constituting the most 5′-contig member was taken as the representative for the unigene set. The unigene set was then picked into new 96-well microtitre plates. After picking, the accuracy of the unigene set was confirmed by re-sequencing a 15%, evenly distributed, subsample of the picked clones. No discrepancies with clone identity or expected position were identified during this re-sequencing process. The cDNA inserts from each of the unigene clones were polymerase chain reaction (PCR) amplified and purified as described in the supplementary data. Electrophoretic assessment of 20% of the amplified products gave an approximate PCR success rate for the unigene set of 96% (data not shown), with success being defined as the production of a single PCR product with no residual primers and a final concentration in the spotting solution in excess of 200 ng/μL.

Unigene microarrays and array hybridization experiments

Microarrays of 19 846 spots were arrayed on to Amersham Codelink activated slides with adjacent duplicate spotting and the inclusion of some redundancy as an internal control and other positive and negative controls. Positive controls comprised PCR products corresponding to the human spike controls described, whilst the negative controls included PCR products corresponding to other human-specific sequences, the multiple cloning site from pBluescript II SK+, the *Amp*^{r}, *BAR* and *GUS* genes. The positions and identities of all the products arrayed can be viewed as one of two Excel spreadsheets in the supplementary data.

In order to ascertain the quality of the printed microarrays, a number of dye swap hybridization experiments were performed using fluorescent-labelled probes derived from

RNA isolated from embryos at 40 dpa and at 1 day post-germination. Raw data of spot fluorescence intensities were collected from scanned slides and analysed using GenePix 4. Scatter plot analysis of the resulting hybridization signals generated by individual slides (see Figure 3) indicated that 93.9% of spots produced hybridizing signals above background over a wide range of detection sensitivity. Hybridization of reverse dye-labelled probes to duplicate slides was visually judged to be successful in that reverse labelling of spots was obvious (see Figure 4). There was consistently a small degree of intensity-dependent bias towards the fluorescence signal generated by the hybridization of probes containing Alexa Fluor dye 555 conjugation to aminoallyl-labelled cDNAs. Approximately 97.6% of the spots on the array were considered to be good features (more than 55% of their feature pixels brighter than the median background intensity at both scanning wavelengths). Approximately 98.0% of the replicate hybridizing spot pairs were judged to be good features (the standard deviation of the ratio of the median fluorescences between the spots at either or both 532 and 635 nm was less than 0.3).

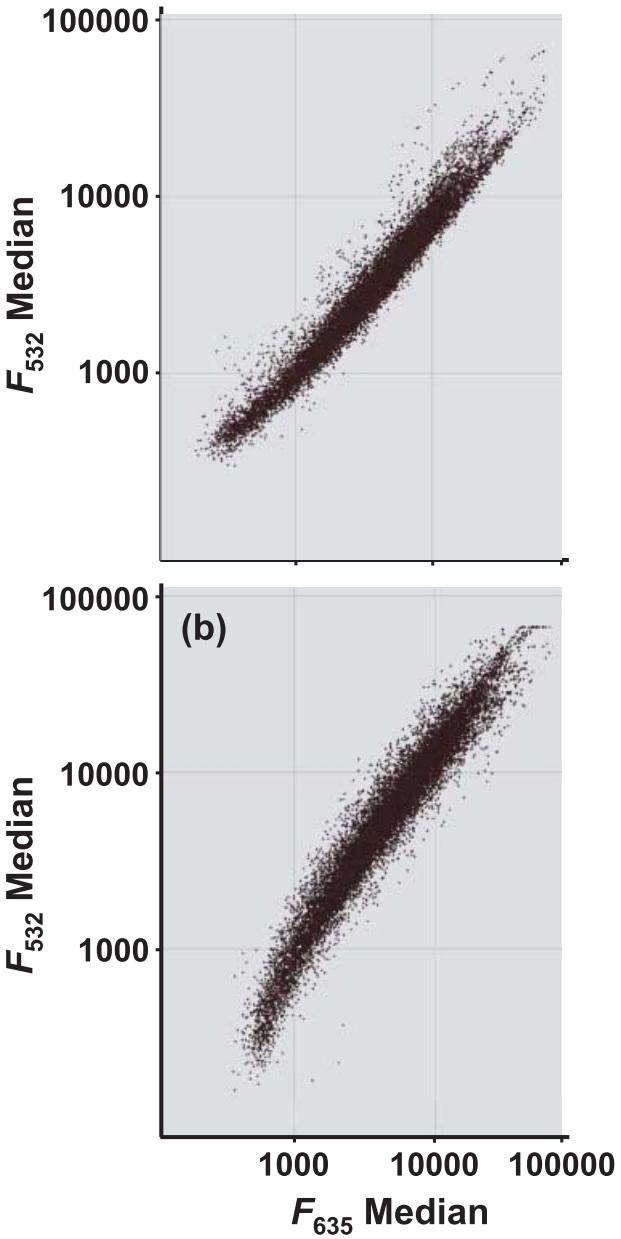

Figure 3 GenePix 4-generated scatter plots showing the data resulting from reverse dye-labelled cDNA microarray hybridization experiments using the wheat unigene arrays. (a) Total RNA from embryos at 40 days post-anthesis (dpa) was used to generate Alexa Fluor 555-conjugated aminoallyl-labelled cDNAs, mixed with a similar amount of Alexa Fluor 647-conjugated aminoallyl-labelled cDNA generated from the total RNA of embryos at 1 day post-germination (dpg) and cohybridized to a wheat unigene microarray. (b) The reverse experiment was performed in which total RNA from embryos at 40 dpa was used to generate Alexa Fluor 647-conjugated aminoallyl-labelled cDNAs, mixed with a similar amount of Alexa Fluor 555-conjugated aminoallyl-labelled cDNAs generated from the total RNA of embryos at 1 dpg and cohybridized to a wheat unigene microarray.

Discussion

Library construction and development of an EST resource

We have produced a publicly available, quality-controlled, hexaploid wheat cDNA library and EST resource of 26 382 sequences derived from 35 individual cDNA libraries relating to a number of highly specific developmental stages of different tissues of both grains and seedlings. In terms of the average insert size and the percentage of non-recombinants present, the individual libraries broadly fell within the generally accepted qualitative boundaries for those produced by the methods described. Some libraries, such as O2:32 (combined aleurone and endosperm tissue 2 days post-germination), proved to be more difficult to construct, mainly because of the scarcity of intact transcripts within the tissues used for RNA extraction. However, even these libraries provided good-quality sequence data. Inevitably, our strict approach has reduced the overall sequencing efficiency of the programme to around 60% and may have inadvertently excluded sequences encoding proteins with molecular weights below approximately 3.5 kDa. However, the sequence length of at least 100 base pairs and the PHRED-determined sequencing accuracy of 99% indicate that these issues were not important. The number of sequences for which no function could be assigned was 25% of the total. This is a very conservative estimate and the use of a lower *E*-value cut-off

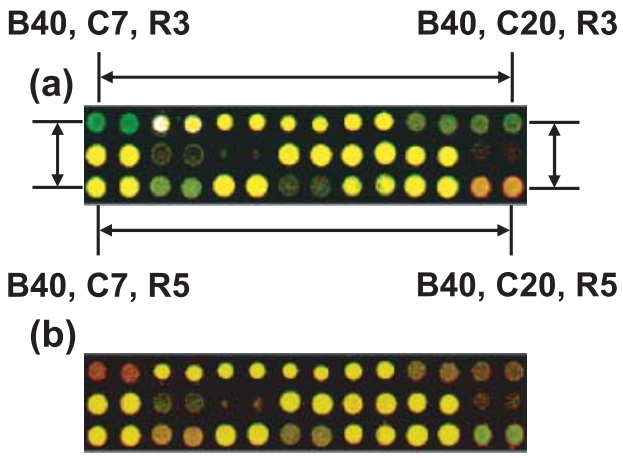

Figure 4 Sections of GenePix 4-generated images of wheat unigene microarrays used in reverse dye-labelled cDNA hybridization experiments. The same section of two individual hybridized arrays is shown in both (a) and (b). In each case, the section shown is from block (B) 40 of the unigene array and is from column (C) 7 to C20 in the horizontal direction and from row (R) 3 to R5 in the vertical direction. (a) Total RNA from embryos at 40 days post-anthesis (dpa) was used to generate Alexa Fluor 555-conjugated aminoallyl-labelled cDNAs, mixed with a similar amount of Alexa Fluor 647-conjugated aminoallyl-labelled cDNA generated from the total RNA of embryos at 1 day post-germination (dpg) and cohybridized to a wheat unigene microarray. (b) The reverse experiment was performed in which total RNA from embryos at 40 dpa was used to generate Alexa Fluor 647-conjugated aminoallyl-labelled cDNAs, mixed with a similar amount of Alexa Fluor 555-conjugated aminoallyl-labelled cDNAs generated from the total RNA of embryos at 1 dpg and cohybridized to a wheat unigene microarray.

so as to include a greater number of sequences of potentially doubtful functional similarity could have been used to provide a more realistic figure of around 40%. For a number of sequences (10%) there was no functional similarity using BLASTX; however, BLASTN did reveal potential similarities. In the main, this simply reflected the presence of short 3′-untransformed region (3′-UTR) sequences within the data set, but also pertained to the presence of a very few sequences with multiple frame shifts that interrupt correct translation and a small (0.67%) contamination with rRNA gene sequences. Approximately 5.5% of the sequences resulting from this programme had been defined as unique in that no similarity to any pre-existing sequences could be found using either BLASTX or BLASTN to search against either the non-redundant or EST data sets held at the NCBI. The current cereal unigene sets (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene) are built, respectively, from the clustering of 390 956 *Triticum aestivum*, 300 902 *Hordeum vulgare*, 291 667 *Zea mays*, 260 575 *Oryza sativa* and 107 607 *Sorghum bicolor* cDNA and EST sequences. With the additional abundance of cDNA and EST sequences available for *A. thaliana* and other species, this further addition of 1450 unique wheat sequences into the public domain was surprising. This may have reflected the overall strategy employed here, in that a greater proportion of novel sequences may have been obtained by sequencing at less depth from a larger number of highly specific and developmentally diverse libraries than by sequencing at greater depth from fewer developmentally similar libraries. Certainly this would be the case where a few sequences represented a large percentage of the total in one given tissue type.

The careful tagging of ears at anthesis and the detailed microscopic examination of the harvested tissues (see http://www.wheatbp.net/) have enabled the EST sequences described here to be developmentally assigned with a high degree of confidence. This has provided a higher level of confidence for the development of virtual Northern and differential display (see http://www.cerealsdb.uk.net/database.htm) approaches for the discovery of novel candidate genes. This contrasts with the application of these mining approaches to wheat sequence data sets held in public databases, where difficulties can arise as a result of poor sequence annotation. For example, comparing the libraries P1:33 and P6:38, it was possible to show that the cerealsdb clone E12_p133_plate_14 (acc. no. AL821594), which encodes a dehydrin, was expressed at significantly ($P = 0.01$, Z score = 5.91) higher levels in seedlings in response to cold treatment, and that the mRNA encoding the storage protein γ-gliadin was expressed during the later rather than the early stages of endosperm development. Such precision is impossible where the annotation accompanying sequences within public databases is less precise. For example, EST sequences identical to that encoding the above-mentioned dehydrin are, amongst other entries, publicly described as originating from cDNA libraries made from tissues isolated from an unrecorded developmental stage of the lemma and palea, from anthers at an undefined stage of development, from a mixture of both cold-treated and salt-stressed roots and from the adult plant at some stage prior to anthesis. The situation is similar for a very large number of the wheat sequences within the public databases.

Development of a wheat EST microarray resource

Gaining an understanding of the function of the thousands of genes in the cereal genome is one of the ultimate goals of functional genomics. The use of microarrays to investigate global alterations in the transcriptome during tissue development or following changes in the environment offers one correlative approach by which biological function can be rapidly assigned. For some plants, such as *A. thaliana* and barley, microarray chips comprising up to 10 000 genes are commercially available (http://www.affymetrix.com/). A number

of publicly funded programmes, such as AFGC (Arabidopsis Functional Genomics Consortium; http://afgc.stanford.edu) and GARNet (Genomic Arabidopsis Resource Network; http://www.york.ac.uk/res/garnet/may.htm), offer services where microarrays can be hybridized to fluorescent-labelled probes derived from RNA samples of choice. Many aspects of the development and biology of *A. thaliana* have been examined using microarrays to identify associated changes in the transcriptome. For example, recent studies have analysed alterations in the transcriptome of *A. thaliana* in response to herbicide treatment (Lechelt-Kunze *et al*., 2003), pathogen interactions (Tao *et al*., 2003; Whitham *et al*., 2003), altered brassinosteroid levels (Mussig *et al*., 2002) and during the cell cycle (Menges *et al*., 2002). However, until now there have been limited opportunities to perform similar experiments with wheat. In particular, work to identify the factors that determine the nutritional and bread making quality of the grain has been limited by an inability to study more than a few genes simultaneously. The contribution made by the gluten proteins, the gliadins and glutenins, in terms of their roles in determining end use quality, has been reviewed extensively (Shewry *et al*., 2003a, 2003b) and the expression of many of the genes encoding these proteins has been determined (Clarke *et al*., 2002, 2003). Similarly, starch synthesis in the cereal endosperm and the impact of starch structure on quality have attracted much attention (Shewry and Morell, 2001; Burrell, 2003; James *et al*., 2003). Mapped QTLs (Quantitative Trait Loci) associated with many other important aspects of cereal performance and development are also well documented and available for online querying using the GrainGenes database (http://wheat.pw.usda.gov/index.shtml) (Matthews *et al*., 2003). Some of these have been reviewed recently (Appels *et al*., 2003). However, although some 551 000 ESTs for wheat are currently publicly available, there are few known associations of these with either QTLs or aspects of wheat growth and development. We believe that, by generating the wheat microarray resource described here, it should be possible for researchers to identify further associations. Hence, we and a number of other groups are currently using these microarrays in numerous projects covering areas such as the transition from dormancy to germination, nitrogen use efficiency, genotype–environment interaction and the influence that transgenes have on the expression profile of the developing grain. We hope to report on all of these projects in the near future.

Experimental procedures

Below is a concise account of the procedures used in this study. Full details of all the experiments performed can be found in the supplementary data.

Plant material

Except where stated, all experiments were performed using European hexaploid winter wheat (var. Mercia). All plant growth experiments were replicated twice and samplings at each time point were performed in duplicate. Developed inflorescences were date tagged at the first sign of anther emergence and the first spikelets containing flowers undergoing anthesis in individual ears were also noted. Experimental tissues were then sampled at defined post-anthesis time points from the developing grain of the two lower florets of these spikelets and were identified and microdissected under magnification using a Leica MZ6 binocular microscope (Leica Microsystems UK Ltd, Milton Keynes, Buckinghamshire, UK). For carpel tissue at 1 and 2 dpa, whole seeds were carefully 'rolled out' of the floret intact, contaminating pollen was brushed away using a fine paintbrush and the grains were immediately frozen in liquid N_2. At 6, 8, 10 and 14 dpa, the embryos from the developing grains were isolated as enriched tissue fractions. Thin slices containing the developing embryos were excised from the grains and immediately frozen in liquid N_2. At 21, 28, 30 and 40 dpa, the much more obvious embryos were simply detached whole from the grains using a scalpel and were similarly frozen. At 6 dpa the liquid endosperm of the grain was collected by halving the grain and rapidly washing out the endosperm cavity with phosphate-buffered saline (PBS). The PBS containing the liquid endosperm material was then immediately frozen on to the surface of glass microscope slides placed on dry ice. At 8, 10 and 14 dpa, the soft endosperm tissue was simply squeezed from the grains after removal of the embryo region and was directly frozen in liquid N_2. At 21, 28, 30 and 40 dpa, the much more starchy endosperm tissue was isolated from halved grains using a scalpel to scrape the cells directly into liquid N_2. At each stage the residual maternal tissue was frozen in liquid N_2 following removal of the embryo and endosperm tissues. The frozen, dissected tissues were stored at –70 °C prior to RNA extraction. Various whole seedling stress experiments were performed in controlled environment chambers using hydroponic culture techniques (Hewitt, 1966). Whole seedling cold acclimation/stress experiments were performed in controlled environment chambers using 3-week-old seedlings grown in vermiculite soaked in culture solution.

Photographic imaging of experimental material

Whole and cut wheat grain tissues were fixed and embedded in Lambwax (Raymond A. Lamb Ltd, Eastbourne, UK) using standard histochemical techniques (Jensen, 1962) described

in the supplementary data. Slides were examined at low magnification under bright-field illumination using a Wild Photomacroscope IV (Leica Microsystems UK Ltd). A Leica DMRB microscope (Leica Microsystems UK Ltd) was used for high-magnification observations under similar illumination. Images were recorded on Kodak EPT film (Kodak Ltd, Hemel Hempstead, UK).

Total RNA extractions and polyA+ mRNA purification

Total RNA was extracted from each of the tissues using the method described by Chen *et al*. (2001), except that the extraction buffer was replaced by a sodium dodecylsulphate (SDS)/phenol buffer and much of the cell material complexed with the SDS was removed from the homogenate before phenol/$CHCl_3$ extraction by the addition of 0.2 vol of 5 M KAc and centrifugation. PolyA+ mRNA was purified from total RNA samples by oligo$(dT)_{18–30}$ cellulose (Stratagene, La Jolla, CA, USA) column chromatography, as described by Sambrook *et al*. (1989), except that the lauryl sarcosine was omitted from the solutions applied to the columns.

cDNA synthesis, library construction, quality assessment, colony picking and library storage

Oligo(dT)-linker-primed, hemimethylated, 5′-*Eco*RI-cDNA-*Xho*1-3′ was synthesized from 5 μg aliquots of polyA+ mRNA and size-fractionated using a commercially available cDNA synthesis kit (Stratagene), as instructed by the manufacturer. Size-fractionated cDNA samples were chosen for ligation if the cDNA exhibited a minimum size of approximately 500 base pairs with an average size of approximately 1.5 kilobase pairs. Aliquots of the prepared, size-fractionated cDNA were directionally ligated with 50 ng of calf, intestinal, alkaline phosphatase (CIAP)-treated, *Eco*RI/*Xho*1-digested pBluescript II SK+ (Stratagene) at a cDNA : vector ratio of 2 : 1 using T4 DNA ligase (Stratagene). Each ligation was then used in five replicate transformations of *E. coli* XL10 Gold ultracompetent cells (Stratagene) as instructed by the manufacturer. Aliquots of the transformed *E. coli* cells were plated on to 22 cm × 22 cm Luria–Bertani (LB) agar plates containing 75 μg/mL ampicillin and 12.5 μg/mL tetracyclin. Following positive quality assessment, a Q-Bot (Genetix Ltd, Queensway, New Milton, Hampshire, UK) automated picker was used to pick colonies from each library made into 96-well microtitre plates (Abgene, Epsom, UK) containing 100 μL aliquots of LB medium with 75 μg/mL ampicillin and 12.5 μg/mL tetracyclin. Following incubation to allow for cell growth, sterile glycerol was added to 25% (v/v) to each of the wells and the plates were then placed and stored at –70 °C prior to sequencing.

Sequencing template preparation and high-throughput sequencing of cDNA inserts

EST DNA sequencing templates were prepared for high-throughput sequencing via a colony PCR (Saiki *et al*., 1988) approach in 96-well format. Following the PCR amplification of template DNA, sequencing was performed using a dye terminator cycle sequencing approach. Individual sequences were read in 96-well format using a MegaBACE 1000 (Amersham Biosciences UK Ltd, UK) capillary sequencing machine, as directed by the manufacturer. Sequence data were output in the form of MegaBACE format ESD files.

Bioinformatics and sequence data handling

EST sequence cleanup and BLAST searching were handled by a custom Laboratory Information Management (LIM) system using PERL (http://www.perl.org) and MySQL (http://www.mysql.com) on a Redhat Linux 7 (http://www.redhat.com) PC. MegaBACE ESD format files were base called and quality trimmed using PHRED (Ewing *et al*., 1998) with default trim_alt parameters. pBluescript II SK+ vector, library linker-primer and *Eco*RI adapter sequences were masked using Cross_match©, an efficient implementation of the Smith–Waterman (Smith and Waterman, 1981) sequence alignment algorithm. Custom PERL scripts were subsequently used to select the longest unmasked region of each sequence and to remove ESTs comprising only low complexity sequence. Sequences were clustered using cap3 (Huang and Madan, 1999) and cluster information was parsed into the database for subsequent analysis and for use in the selection of individual clones for inclusion in a unigene set. Web-based data mining tools for electronic Northerns and digital differential displays were written in PERL-CGI using the PERL-MySQL database interface: these tools are available on request. The EST sequences can be downloaded from http://www.cerealsdb.uk.net or from the EMBL/GENBANK/DDJB public databases (accession numbers AL808219 to AL831324 and AJ601557 to AJ604482).

For microarray experiments the raw data of spot fluorescence intensities were collected from scanned slides into GenePix 4 (Axon Instruments, Inc., Union City, CA, USA). At this stage the slide background fluorescence signal was removed from the spot values and an analysis was made of slide quality. The reproducibility of the data within and between slides was examined by performing Pearson correlations of

the red and green channels. Slides with correlation factors of below 0.96 were rejected from further analysis.

Preparation of unigene array templates and microarray production

A unigene set was picked into 96-well format from the original library stock plates using a Qbot (Genetix Ltd) colony picker and following the instructions of the manufacturer. The clone identities of the picked unigene set were confirmed by re-sequencing as before. Unigene set DNA templates for microarraying were produced by 96-well format PCR in the presence of 5′-amino-C6-labelled T3 and T7 oligonucleotides (Sigma Genosys Ltd, Sigma-Aldrich House, Haverhill, Cambridge, UK). After ethanol precipitation, centrifugation and drying, the resulting PCR products were dissolved in 150 mM sodium phosphate pH 8.5 spotting solution and transferred into 384-well format for microarray printing. Duplicate-spotted unigene set arrays (48 blocks of 20 × 21 spots in three fields of 16 blocks arrayed in 4 × 4 layout; see supplementary data) of the PCR products were then printed from 384-well source plates on to Codelink activated microarray slides (Amersham Biosciences UK Ltd) using a BioRobotics Microgrid II arrayer (Genomic Solutions Ltd, Huntingdon, Cambridge, UK) with MicroSpot 2500 split pins, as instructed by the manufacturer. Immediately after spotting, the arrayed slides were maintained at 21 °C in an air-tight chamber above a saturated NaCl solution for 24 h. The slides were then stored at room temperature in a dust-free, desiccated environment before use.

Production of polyA+ mRNA spiking control transcripts

A number of polyA+ cDNAs corresponding to mammalian specifically expressed mRNA transcripts were obtained as plasmid DNAs in the vector pT7T3D-Pac from Dr Virginia Walbot (Stanford University, Stanford, CA, USA) as freely distributable clones from the Maize Gene Discovery Project (http://www.zmdb.iastate.edu/). These included the clones pHUM2 (GB:AA418251), pHUM4 (GB:AA464627), pHUM5 (GB:H28469) and pHUM8 (GB:AA485668). Also, a full-length, 2403 base pair, polyA+ cDNA encoding human tumour necrosis factor receptor associating factor 1 (TRAF1) (GB:BC024145) was obtained from the MRC geneservice (Babraham, Cambridge, UK) as an *Eco*RI/*Xho*1-cloned fragment in the vector pOTB7. Initially, using standard approaches, the *Eco*RI/*Xho*1 TRAF1-encoding fragment was subcloned into similarly digested pBSII SK+ to yield the clone pTRAF1. After *Not*1 digestion of plasmid DNAs, mRNA transcripts were subsequently produced *in vitro* from pHUM2 and pHUM4 using T7 RNA polymerase and from *Xho*1-digested pTRAF1 using T3 RNA polymerase. Transcript control mixes were prepared in SDW (sterile distilled water) such that 3 μL contained 1 ng of pHUM2, 0.1 ng of pHUM4 and 0.05 ng of pTRAF1 polyA+ mRNA transcripts combined.

Microarray probe generation and array hybridizations

Microarray slides were hybridized in reverse dye labelling experiments using Alexa Fluor dye 555 and 647 (Molecular Probes Inc., Eugene, OR, USA) secondarily conjugated, aminoallyl-labelled cDNA probes generated from total RNA samples. After spiking with different control mRNA transcripts at various individual concentrations, total RNA samples were reverse transcribed with Stratascript reverse transcriptase (Stratagene) in the presence of 5-(3-aminoallyl)-2′-deoxyuridine 5′-triphosphate (AA-dUTP) (Sigma) in order to produce aminoallyl-labelled first-strand cDNAs (AA-cDNA). AA-cDNA samples were divided equally for esterification to each of the Alexa Fluor dyes. Succinimidyl esters of the Alexa Fluor dyes were obtained from Molecular Probes as ready to use Alexa Fluor 555 and 647 reactive dye decapacks and were used to label AA-cDNA samples as instructed. Following labelling, the fluor dye-esterified AA-cDNA samples were again purified using MinElute PCR purification spin columns. The eluted, labelled cDNA samples were then used immediately in array hybridization experiments.

Arrays were pretreated and hybridized to fluor dye-labelled AA-cDNA samples as instructed by the slide manufacturer. After hybridization, the arrays were washed successively for 5 min in 2 × SSC, 0.1% (w/v) SDS at 42 °C, 5 min in 1 × SSC, 0.1% (w/v) SDS at 42 °C, 1 min in 0.2 × SSC, 0.1% (w/v) SDS at room temperature, 1 min in 0.1 × SSC at room temperature and 0.5 min in 100% (v/v) ethanol at room temperature. The array slides were then immediately dried by centrifugation and the resulting hybridization was visualized by scanning using an Axon Instruments GenePix 4000B dual laser microarray scanner with integrated GenePix 4 software for data acquisition.

Acknowledgements

We are grateful to the Biotechnology and Biological Sciences Research Council, UK (BBSRC), for providing the main funding for this work under the Cereals 'Investigating Gene Function' initiative (ref. IGF12403). Chungui Lu was funded by a BBSRC Agri-Food award (ref. D16781) and Rebecca Lyons by a BBSRC Gene Flow Agri-Food special initiative (ref. GM114152). We are also grateful to Professor Graham Jellis

at the Home Grown Cereals Authority for providing funding to Jill Parker for the development of the online micrographic resource (http://www.wheatbp.net/) detailing wheat grain development. Our thanks to Bob Hughes, Richard Parkinson and Derek Edgell for their invaluable assistance with the controlled environment growth of wheat plants and to Les Saker for his help with hydroponic systems. Our thanks also to Dr Robbie Waugh and Dr David Caldwell of the Scottish Crops Research Institute, UK, for their assistance in helping pick the unigene set, and to Dr Graham Moore of the John Innes Centre, UK, for coordinating the IGF programme. This article is dedicated to the memory of the now closed Long Ashton Research Station.

Supplementary material

The authors have supplied detailed experimental procedures, which are available as a supplementary Appendix from http://www.blackwellpublishing.com/products/journals/suppmat/PBI/PBI096/PBI096sm.htm.

References

Aharoni, A., Keizer, L.C., Bouwmeester, H.J., Sun, Z., Alvarez-Huerta, M., Verhoeven, H.A., Blaas, J., van Houwelingen, A.M., De Vos, R.C., van der Voet, H., Jansen, R.C., Guis, M., Mol, J., Davis, R.W., Schena, M., van Tunen, A.J. and O'Connell, A.P. (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell*, **12** (5), 647–662.

Appels, R., Francki, M. and Chibbar, R. (2003) Advances in cereal functional genomics. *Funct. Integr. Genomics*, **3** (1–2), 1–24.

Aquino, P., Carron, F. and Calvo, R. (1999) Selected wheat statistics. In *CIMMYT 1998–99 World Wheat Facts and Trends. Global Wheat Research in a Changing World: Challenges and Achievements* (Pingali, P.L., ed.), pp. 33–45. Mexico, D.F.: CIMMYT.

Burrell, M.M. (2003) Starch: the need for improved quality or quantity – an overview. *J. Exp. Bot*. **54** (382), 451–456.

Chen, 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** (5–6), 799–807.

Clarke, B.C., Hobbs, M., Skylas, D. and Appels, R. (2002) Genes active in developing wheat endosperm. *Funct. Integr. Genomics*, **1** (1), 44–55.

Clarke, B.C., Phongkham, T., Gianibelli, M.C., Beasley, H. and Bekes, F. (2003) The characterisation and mapping of a family of LMW-gliadin genes: effects on dough properties and bread volume. *Theor. Appl. Genet*. **106** (4), 629–635.

Ewing, B., Hillier, L., Wendl, M.C. and Green, P. (1998) Base-calling of automated sequencer trace files using phred. I. Accuracy assessment. *Gen. Res*. **8**, 175–185.

Girke, T., Todd, J., Ruuska, S., White, J., Benning, C. and Ohlrogge, J. (2000) Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol*. **124** (4), 1570–1581.

Hewitt, E.J. (1966) *Sand and Water Culture Methods Used in Sand and Water Culture. Technical Communication*, 2nd edn. East Malling, Kent: Commonwealth Bureau of Horticulture and Plantation Crops.

Holtorf, H., Guitton, M.C. and Reski, R. (2002) Plant functional genomics. *Naturwissenschaften*, **89** (6), 235–249.

Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Gen. Res*. **9**, 868–877.

James, M.G., Denyer, K. and Myers, A.M. (2003) Starch synthesis in the cereal endosperm. *Curr. Opin. Plant Biol*. **6** (3), 215–222.

Jensen, W.A. (1962) *Botanical Histochemistry: Principles and Practice*. San Francisco, CA: W.H. Freeman.

Kawasaki, S., Borchert, C., Deyholos, M., Wang, H., Brazille, S., Kawai, K., Galbraith, D. and Bohnert, H.J. (2001) Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell*, **13** (4), 889–905.

Kollipara, K.P., Saab, I.N., Wych, R.D., Lauer, M.J. and Singletary, G.W. (2002) Expression profiling of reciprocal maize hybrids divergent for cold germination and desiccation tolerance. *Plant Physiol*. **129** (3), 974–992.

Langridge, P., Lagudah, E.S., Holton, T.A., Appel, R., Sharp, P.J. and Chalmers, K.J. (2001) Trends in genetic and genome analysis in wheat: a review. *Aust. J. Agric. Res*. **52**, 1043–1077.

Lechelt-Kunze, C., Meissner, R.C., Drewes, M. and Tietjen, K. (2003) Flufenacet herbicide treatment phenocopies the fiddlehead mutant in *Arabidopsis thaliana*. *Pest Manag. Sci*. **59** (8), 847–856.

Lee, J.M., Williams, M.E., Tingey, S.V. and Rafalski, J.A. (2002) DNA array profiling of gene expression changes during maize embryo development. *Funct. Integr. Genomics*, **2** (1–2), 13–27.

Lockhart, D.J. and Winzeler, E.A. (2000) Genomics, gene expression and DNA arrays. *Nature*, **405** (6788), 827–836.

Matthews, D.E., Carollo, V.L., Lazo, G.R. and Anderson, O.D. (2003) GrainGenes, the genome database for small-grain crops. *Nucleic Acids Res*. **31** (1), 183–186.

Menges, M., Hennig, L., Gruissem, W. and Murray, J.A. (2002) Cell cycle-regulated gene expression in *Arabidopsis*. *J. Biol. Chem*. **277** (44), 41 987–42 002.

Mussig, C., Fischer, S. and Altmann, T. (2002) Brassinosteroid-regulated gene expression. *Plant Physiol*. **129** (3), 1241–1251.

Narusaka, Y., Narusaka, M., Seki, M., Ishida, J., Nakashima, M., Kamiya, A., Enju, A., Sakurai, T., Satoh, M., Kobayashi, M., Tosa, Y., Park, P. and Shinozaki, K. (2003) The cDNA microarray analysis using an *Arabidopsis* pad3 mutant reveals the expression profiles and classification of genes induced by *Alternaria brassicicola* attack. *Plant Cell Physiol*. **44** (4), 377–387.

Negishi, T., Nakanishi, H., Yazaki, J., Kishimoto, N., Fujii, F., Shimbo, K., Yamamoto, K., Sakata, K., Sasaki, T., Kikuchi, S., Mori, S. and Nishizawa, N.K. (2002) cDNA microarray analysis of gene expression during Fe-deficiency stress in barley suggests that polar transport of vesicles is implicated in phytosiderophore secretion in Fe-deficient barley roots. *Plant J*. **30** (1), 83–94.

Ogawa, M., Hanada, A., Yamauchi, Y., Kuwahara, A., Kamiya, Y. and Yamaguchi, S. (2003) Gibberellin biosynthesis and response during *Arabidopsis* seed germination. *Plant Cell*, **15** (7), 1591–1604.

Ruan, Y., Gilmore, J. and Conner, T. (1998) Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. *Plant J*. **15** (6), 821–833.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T.,

Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Shewry, P.R., Halford, N.G. and Lafiandra, D. (2003b) The genetics of wheat gluten proteins. *Adv. Genet*. **49**, 111–184.

Shewry, P.R., Halford, N.G., Tatham, A.S., Popineau, Y., Lafiandra, D. and Belton, P.S. (2003a) The high molecular weight subunits of wheat glutenin and their role in determining wheat processing properties. *Adv. Food Nutrition Res*. **45**, 221–302.

Shewry, P.R. and Morell, M. (2001) Manipulating cereal endosperm structure, development and composition to improve end-use properties. *Adv. Bot. Res*. **34**, 165–236.

Smith, T.F. and Waterman, M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol*. **147**, 195–197.

Tao, Y., Xie, Z., Chen, W., Glazebrook, J., Chang, H.S., Han, B., Zhu, T., Zou, G. and Katagiri, F. (2003) Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell*, **15** (2), 317–330.

Wang, R., Guegler, K., LaBrie, S.T. and Crawford, N.M. (2000) Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *Plant Cell*, **12** (8), 1491–1509.

Wang, R., Okamoto, M., Xing, X. and Crawford, N.M. (2003) Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol*. **132** (2), 556–567.

Whitham, S.A., Quan, S., Chang, H.S., Cooper, B., Estes, B., Zhu, T., Wang, X. and Hou, Y.M. (2003) Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. *Plant J*. **33** (2), 271–283.

Wilson, I.D., Barker, G.L. and Edwards, K.J. (2003) Genotype to phenotype: a technological challenge. *Ann. Appl. Biol*. **142**, 33–39.

Yu, L.X. and Setter, T.L. (2003) Comparative transcriptional profiling of placenta and endosperm in developing maize kernels in response to water deficit. *Plant Physiol*. **131** (2), 568–582.