Plant Biotechnology Journal

Plant Biotechnology Journal (2009) **7**, pp. 401–410 doi: 10.1111/j.1467-7652.2009.00405.x

A novel transcriptomic approach to identify candidate genes for grain quality traits in wheat

Yongfang Wan[1], Claudia Underwood[1], Geraldine Toole[2], Peter Skeggs[3], Tong Zhu[4], Michelle Leverington[5], Simon Griffiths[5], Tim Wheeler[6], Mike Gooding[6], Rebecca Poole[6], Keith J. Edwards[7], Salvador Gezan[1], Sue Welham[1], John Snape[5], E. N. Clare Mills[2], Rowan A. C. Mitchell[1] and Peter R. Shewry[1,*]

[1]*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*
[2]*Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK*
[3]*Hovis, part of Premier Foods, Lord Rank Centre, Lincoln Road, High Wycombe, Buckinghamshire HP12 3QR, UK*
[4]*Syngenta Biotechnology Inc., PO Box 12257, Research Triangle Park, NC 27709-2257,USA*
[5]*John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK*
[6]*University of Reading, Earley Gate, PO Box 236, Reading, Berkshire RG6 6AR, UK*
[7]*University of Bristol, Biological Sciences Building, Woodland Road, Clifton, Bristol BS8 1UG, UK*

Received 22 September 2008;
revised 28 November 2008;
accepted 8 December 2008.
**Correspondence* (fax +44 (0)1582 763010;
email peter.shewry@bbsrc.ac.uk)

Keywords: breadmaking, grain quality, transcriptome, wheat.

Summary

A novel methodology is described in which transcriptomics is combined with the measurement of bread-making quality and other agronomic traits for wheat genotypes grown in different environments (wet and cool or hot and dry conditions) to identify transcripts associated with these traits. Seven doubled haploid lines from the Spark × Rialto mapping population were selected to be matched for development and known alleles affecting quality. These were grown in polytunnels with different environments applied 14 days post-anthesis, and the whole experiment was repeated over 2 years. Transcriptomics using the wheat Affymetrix chip was carried out on whole caryopsis samples at two stages during grain filling. Transcript abundance was correlated with the traits for approximately 400 transcripts. About 30 of these were selected as being of most interest, and markers were derived from them and mapped using the population. Expression was identified as being under *cis* control for 11 of these and under *trans* control for 18. These transcripts are candidates for involvement in the biological processes which underlie genotypic variation in these traits.

Introduction

It has been estimated that wheat currently feeds about 35% of the Earth's population, being grown on over 205 million hectares with a total annual harvest of over 600 million tonnes (Food and Agriculture Organization: http://www.fao.org/statistics/). It is only consumed by humans after processing into bread, noodles and other foods, and so processing quality is an important target for breeders.

Substantial genetic variation exists in grain quality, and breeders routinely select for 'high-quality' protein alleles in their breeding programmes (reviewed by Payne, 1987; Shewry *et al*., 2003; Cornish *et al*., 2006). However, there are also significant environmental and genotype × environment (G × E) effects which limit the stability of grain processing characteristics and are poorly understood. Furthermore, previous studies of environmental effects have largely focused on the effects of high temperatures, when the heat shock response is activated (reviewed by Blumenthal *et al*., 1993; Dupont and Altenbach, 2003), conditions which are less relevant to the cool temperate climate in Western Europe. The stability of crop yield and quality to environmental fluctuations will also become more important as climate change is predicted to result in increased year-to-year variation, as well as long-term changes in the climate (Porter and Semenov, 2005; Richter and Semenov, 2005).

Table 1 Characteristics of the seven selected doubled haploid (DH) lines from the Spark × Rialto cross, showing high molecular weight (HMW) subunit alleles and alleles at four quality quantitative trait loci (QTLs) (S, Spark; R, Rialto). The allele confering good quality is indicated

		HMW subunit alleles			QTLs			
Line	Near-infrared	1A	1B	1D	3A Loaf volume and height S = good	3B Loaf height S = good	4B Loaf volume R = good Dough stability S = good	6A Dough stability and development S = good
SR41	R	1	7 + 8	5 + 10	S	R	R	R
SR3	R	Null	7 + 8	5 + 10	S	R	R	R
SR7	S/R	Null	7 + 8	5 + 10	R	S	R	S
SR92	R	Null	7 + 8	5 + 10	R	R	S	S
SR5	S/R	Null	7 + 8	5 + 10	S	R	R	S
SR13	S/R	Null	7 + 8	5 + 10	S	S	S	S
SR107	S/R	Null	7 + 8	5 + 10	S	S	R	S

Includes unpublished data on QTLs from J. Snape *et al*.

The inability of breeders to select for stability to fluctuations in climate may be related to two factors. First, it is difficult to design and implement practical screens to select for stability in breeding programmes. Second, it is probable that multiple genes contribute to differences in stability between cultivars, and these are difficult to select without the ability to discover and identify specific alleles which can be characterized and used for molecular marker-assisted selection.

It is clear that the stability of quality in wheat is a complex trait under multigenic control, and is therefore not amenable to simple mapping approaches. We therefore adopted a new strategy to enable the identification of genes contributing to stability and to facilitate the development of markers by comparing the transcript profiles of doubled haploid (DH) lines grown under various environmental conditions with the processing properties determined by test milling and baking. This identified a number of candidate genes whose expression was correlated with quality traits which were stable over multiple years and environments.

Results

Characteristics of the DH lines

The Spark × Rialto DH population comprises 144 lines (Snape *et al*., 2007). Test baking of 60 lines selected to lack the 1BL/1RS translocation and grown in a field at the John Innes Centre (Norwich, UK) in 2001 showed wide variation in functional properties, with the loaves ranging from 1231 to 1675 mL in volume, 98 to 133 cm in height and 1 to 6 in texture score (data not shown). Some of this variation was clearly related to allelic differences in high molecular weight (HMW) subunit composition. Spark has the HMW subunit composition 1Ax null, 1Bx7, 1By8, 1Dx5 and 1Dy10, whereas Rialto has the composition 1Ax1, 1Bx17, 1By18, 1Dx5 and 1Dy10. The presence of subunit 1Ax1 relative to the null allele and of subunits 1Bx17 and 1By18 would be expected to confer greater dough strength to Rialto (Payne, 1987), but this effect is offset by the 1BL/1RS translocation which has detrimental effects on quality (Graybosch, 2001). The detrimental effects of the translocation are at least partly a result of the replacement of the gluten proteins [gliadins and low molecular weight (LMW) subunits] encoded by loci on 1BS with the secalin proteins encoded by genes on 1RS.

In order to eliminate differences in functional properties as a result of known differences in protein composition, we selected DH lines which lacked the translocation and were matched for HMW subunit alleles. From the Mahalanobis distances derived during near-infrared (NIR) spectroscopy, the relative position of the progeny to their parents was seen to agree well with estimates based on their genetic similarity. Using these data, lines which were spectroscopically 'Spark-like' (S), 'Rialto-like' (R) and in a central region distinct from either parent (S/R) were identified (Table 1). Initially eight lines were selected, all with subunits 1Bx7 and 1By8 and with six having the 1Ax null allele and two the 1Ax1 allele. These lines also differed at four quantitative trait loci (QTLs) which determined differences in processing properties (J. Snape *et al*., unpubl. data). Subsequently one of the two lines expressing subunit 1Ax1 was discarded because of an asynchronous flowering time, leaving the seven lines listed in Table 1.

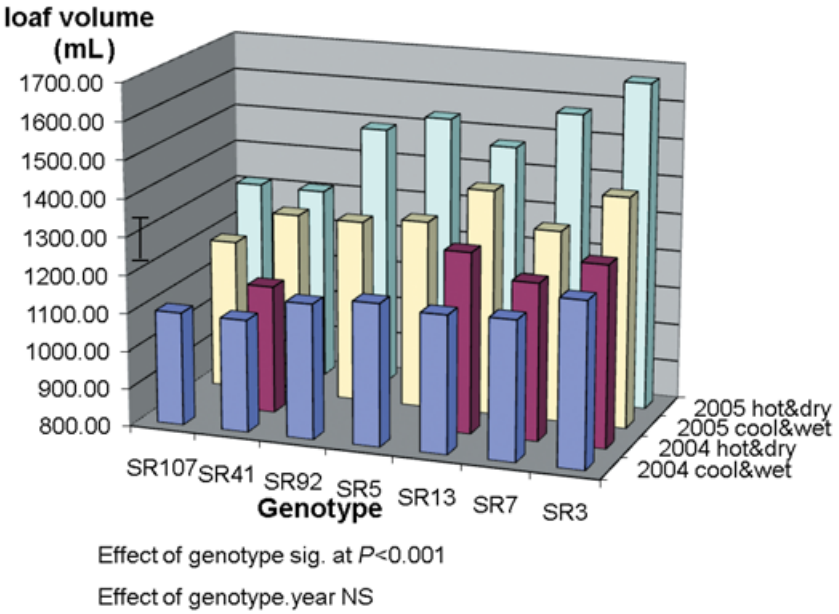

Figure 1 Loaf volumes of the seven Spark × Rialto doubled haploid lines grown in polytunnels in 2004 and 2005, showing consistent genotype effects across environments. No data are available for SR107, SR92 and SR5 grown under hot and dry conditions in 2004 as insufficient grain was produced for milling.

Grain development, yield and quality

Genotype × year × environment means for grain development, yield and quality are described in Tables S2–S11 (see Supporting Information).

Most of the quality parameters showed environmental effects which were attributable largely to changes in the duration of grain filling. In addition, some showed consistent additive effects of genotype, whereas others showed highly significant G × E interactions. In particular, loaf volume, which is probably the single most important quality parameter for breadmaking, showed effects of genotype and environment, but no significant G × E interactions, the ranking of genotypes being consistent with SR3 being the best and SR107 the worst in the four environments (Figure 1). In contrast, the grain sodium dodecylsulphate (SDS) sedimentation volume, which is often used as a rapid indirect measure of potential breadmaking performance, showed significant G × E interactions (Table S4, see Supporting Information).

Transcriptome analysis

In order to identify genotype-dependent transcripts that related to stable differences in functional properties, we decided to combine transcript data for the stages of development, environmental regimes and years into a single dataset for analysis. This comprised data from 56 Affymetrix Genechip® arrays, with single arrays being carried out on each developmental stage × environmental regime × year combination. A set of 1905 probesets from the array was identified as showing significant differences in expression between the seven genotypes. Hierarchical clustering for this set showed completely consistent separation of samples by genotype, and the SR41 line differed most in its expression pattern compared with the other lines (Figure 2).

From this set, a subset of 468 probesets was identified which showed correlation with the genotypic variation in quality parameters (see Experimental procedures). Fifty of these probesets were prioritized as of greatest interest, and 34 of these were mapped successfully within the Spark × Rialto population (J. Snape *et al.*, unpubl. data). Mapping of the transcripts allowed the pattern of alleles among the seven DH lines to be compared with the expression patterns of the transcripts. If the patterns are the same, with either parental allele corresponding to high expression, the allele controlling expression is a *cis* factor. If the patterns differ, it is a *trans* factor. (The probability of a *trans* factor appearing to be *cis* by chance for seven lines is $1/2^6 = 1.6\%$.) The *cis* factors are more informative as the causative sequence is likely to be close to the transcript itself, e.g. the promoter of the gene encoding the transcript.

An example of expression data for three mapped transcripts is shown in Figure 3. These three transcripts had the same pattern of alleles across the seven genotypes, as indicated at the top of the figure. Two of the transcripts showed similar patterns of variation, being associated with high expression (red lines) or low expression (blue lines). This shows that the abundance of these two transcripts was determined by *cis*

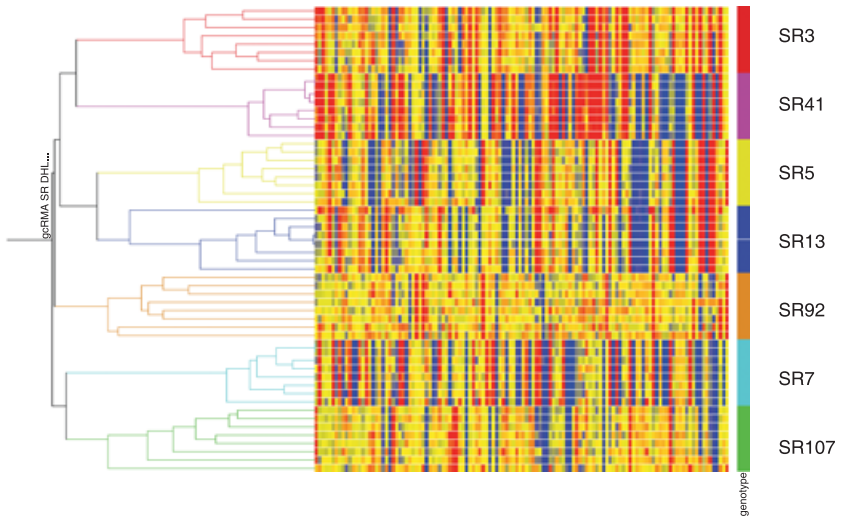

Figure 2 Hierarchical clustering of transcriptome data for the 1905 probesets showing significant genotype-related differences in expression between the seven Spark × Rialto doubled haploid lines.

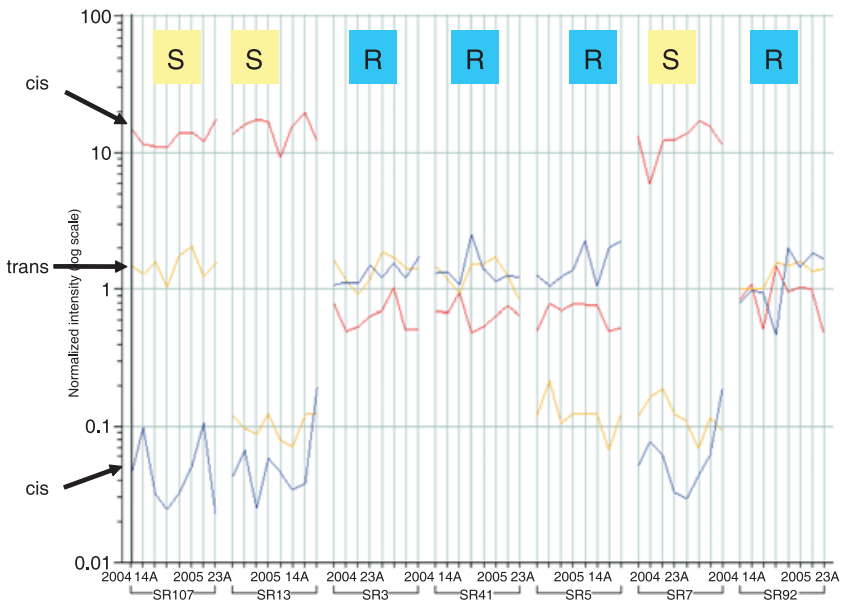

Figure 3 Expression signals for three probesets across 56 arrays. Lines connect the eight values (2 years × 2 stages × 2 tunnels) within each genotype. Markers derived from the transcript sequences were shown to have the same pattern of alleles across the seven genotypes, as indicated at the top (S, Spark allele; R, Rialto allele). The expression patterns indicate that two of the transcripts had *cis* factors controlling expression (red lines, blue lines) and one had a *trans* factor controlling expression (yellow lines).

factors. The third transcript (yellow lines) did not correspond to the expression pattern of the alleles, and the expression of this transcript was therefore controlled by a *trans* factor.

From these analyses, 15 of the mapped transcript sequences were identified as possibly under the control of *cis* factors. The pattern of expression between the seven DH lines and the parents was independently examined using quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) (Figure 4). This supported the hypothesis of *cis* control for 11 of the 15 transcript sequences; the other four exhibited low absolute expression on the Affymetrix chip.

Thus, 11 of the 29 mapped transcript sequences (38%) were shown to be controlled by *cis* factors and 18 (62%) by *trans* factors. The probeset identifiers for these 29 transcripts, the traits with which their expression was correlated and the significance of the correlations are shown in Table 2.

It is possible that some of the apparent differences in expression of the transcripts which appear to be controlled by *cis* factors actually result from sequence polymorphisms which affect the hybridization to the Affymetrix 25-mer probes. As most probes are designed to the 3′ untranslated region of the transcript, such polymorphisms are unlikely to have functional effects. Although such polymorphisms are unlikely to affect the trait of interest, they could nevertheless be useful markers. Inspection of the individual probe signals showed that this explanation would require the presence of

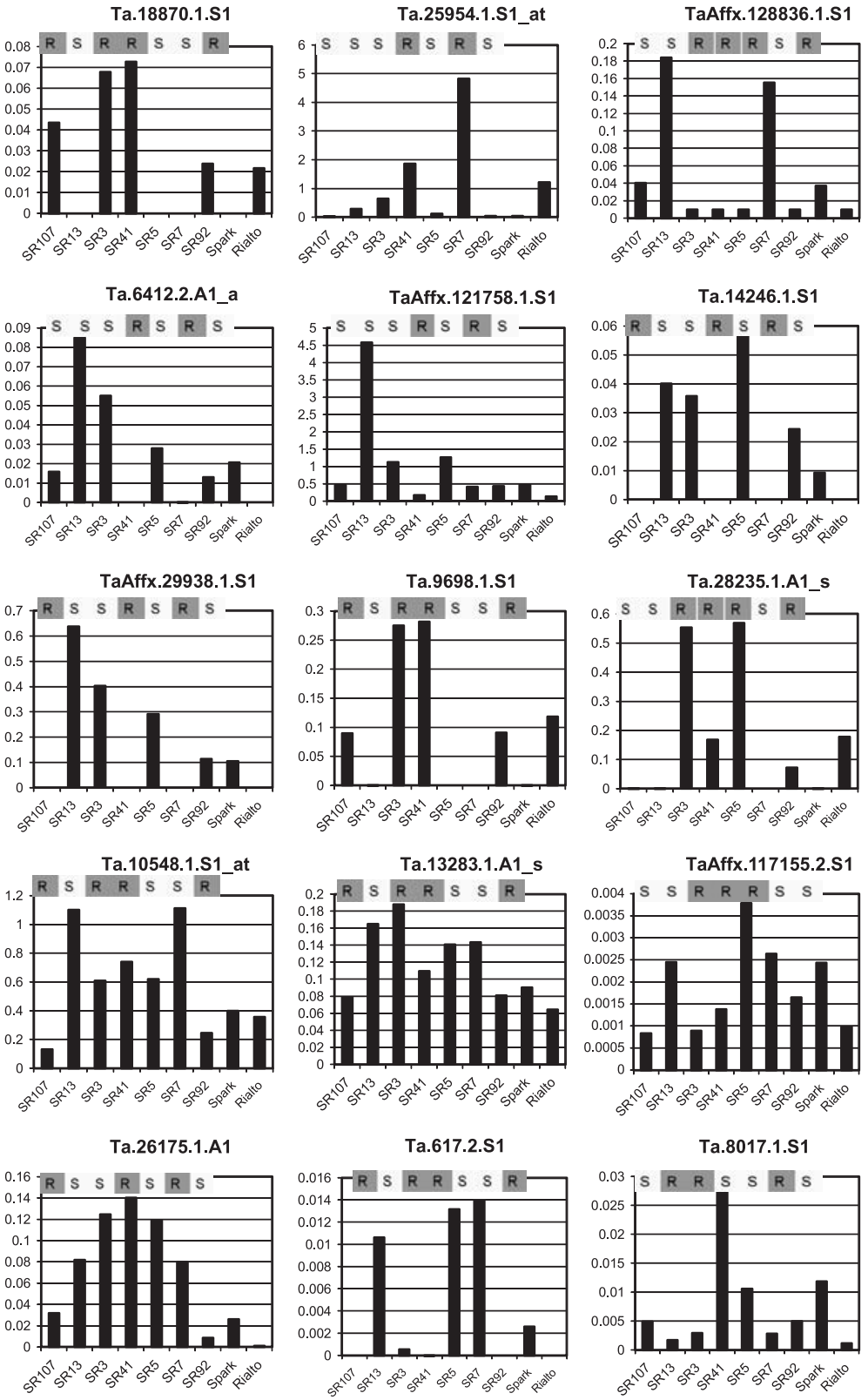

Figure 4 Quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) determination of expression for 15 transcripts identified as probably under *cis* control from Affymetrix data and mapping. Alleles from mapping are indicated. *Cis* control was supported by the qRT-PCR results for all but four of the transcripts; these exceptions were Ta.10548.1, Ta.13283.1, TaAffx.117155.2 and Ta.26175.1.

Table 2 Transcripts showing significant genotype-dependent expression in the seven doubled haploid (DH) lines combining data from 2004 and 2005, both sets of growth conditions (hot and dry, cool and wet) and both developmental stages [14 and 23 days post-anthesis (dpa)]

Identity	Putative function	*cis/trans*	Minimum number SNPs required	resp1	*P*-val1	resp2	*P*-val2	resp3	*P*-val3
TaAffx.44224.1.A1_at	Acetyl transferase involved in gene silencing	*trans*	7	TOTAL_N +	1.8E-04	YIELD15 +	4.3E-03		
Ta.18870.1.S1_at	AP2 transcription factor	*cis*	4	PWA +	8.7E-04	PWA +	1.0E-03	PROT +	3.7E-03
Ta.27780.1.S1_at	β-Amylase-like	*trans*	4	EXTRACT –	7.9E-03	TOTAL_N –	8.2E-03	CSTICK –	8.6E-03
Ta.8640.1.S1_a_at	Oxidoreductase	*trans*	6	GRAIN +	5.5E-04	TOTAL_N +	2.6E-03	TOTAL_S +	3.3E-03
Ta.131.1.S1_at	Low molecular weight glutenin storage protein	*trans*	5	FALLING –	8.5E-04	PWA –	9.5E-04	LOAF_V +	1.3E-03
Ta.9938.1.S1_at	Bifunctional α-amylase	*cis*	8	TOTAL_N +	3.6E-03				
Ta.6984.1.A1_at	PHD finger transcription factor	*trans*	4	PWA +	2.1E-04	FALLING +	2.2E-04	PWA +	4.9E-04
Ta.9223.1.A1_at	Unknown	*trans*	3	FALLING +	1.9E-04	GRAIN +	9.0E-04		
Ta.25954.1.S1_at	Zn finger in ubiquitin-hydrolases	*cis*	3	TOTAL_N +	7.2E-04	YIELD15 +	1.3E-03	SULPHUR –	9.4E-03
Ta.9814.1.S1_at	Zinc finger protein	*trans*	8	LOAF_V +	9.2E-05	PWA –	2.6E-04	LOAF_H +	1.4E-03
Ta.13283.1.A1_s_at	Myb-like DNA-binding	*trans*	6	PWA –	8.9E-05	PWA –	3.5E-03	PROT –	4.9E-03
TaAffx.128836.1.S1_at	Transcription factor X1	*cis*	4	B –	5.6E-05	B –	9.7E-05		
Ta.6412.2.A1_a_at	CBL-interacting protein kinase 9	*cis*	8	TOTAL_N –	9.7E-04	TOTAL_N –	1.8E-03	GRAIN –	3.8E-03
TaAffx.121758.1.S1_at	Phospholipid hydroperoxide glutathione peroxidase	*cis*	6	SULPHUR +	1.2E-04	YIELD15 –	8.8E-04	NITROGEN +	1.2E-03
Ta.25981.1.A1_at	Disease resistance protein	*trans*	10	LOAF_V –	2.7E-04	PWA +	4.4E-04	LOAF_H –	7.7E-04
Ta.14246.1.S1_at	Unknown	*cis*	8	B +	3.4E-06	LOAF_V +	2.0E-03	LOAF_V +	3.2E-03
TaAffx.29938.1.S1_at	Unknown	*cis*	8	B +	1.1E-05	B +	5.3E-05	LOAF_V +	2.8E-03
Ta.9698.1.S1_at	ATP-dependent Clp protease	*trans*	9	PWA +	3.0E-04	PWA +	4.2E-04	PROT +	2.1E-03
Ta.6572.1.S1_at	Peroxiredoxin Q (Prx1)	*trans*	7	GRAIN +	4.2E-04	FALLING +	1.6E-03	GRAIN +	2.4E-03
Ta.28263.1.S1_at	Proline-rich protein	*trans*	5	SULPHUR +	1.7E-03	SULPHUR +	4.4E-03	PROT +	6.5E-03
Ta.25205.1.A1_at	Serine protease	*trans*	6	CSTICK +	3.0E-03	TOTAL_N +	3.3E-03	TOTAL_N +	3.4E-03
Ta.10144.1.S1_at	Histidine-containing phosphotransfer protein	*cis*	7	TOTAL_N +	4.3E-05	TOTAL_S +	1.2E-03	SDS –	5.0E-03
Ta.28235.1.A1_s_at	SHD (SHEPHERD) unfolded protein binding	*trans*	4	B +	6.2E-05	B +	1.1E-04		
Ta.21022.1.S1_at	Proline-rich cell wall protein-like	*trans*	3	FALLING +	4.8E-04	GRAIN +	5.2E-04	GRAIN +	1.2E-03
TaAffx.58820.2.S1_at	Wound-induced protein	*cis*	2	TOTAL_N +	2.2E-03	CSTICK +	3.4E-03	CSTICK +	4.3E-03
Ta.21557.1.A1_at	Senescence/dehydration-associated protein	*trans*	9	TOTAL_N –	1.6E-03	TOTAL_S –	6.1E-03	TOTAL_N –	6.3E-03
Ta.617.2.S1_at	Dolichyl-phosphate β-D-mannosyltransferase	*cis*	3	PWA –	1.0E-04	PROT –	7.9E-03		
Ta.8017.1.S1_at	Embryo-abundant protein EMB	*cis*	2	LOAF_H –	4.3E-03	LOAF_V –	8.0E-03	PWA +	8.2E-03

SNP, single nucleotide polymorphism; resp1/2/3 are the response variates correlated with transcript abundance with significance indicated by P-val1/2/3, respectively. Response variate abbreviations: B flour colour, CSTICK dough stickiness, EXTRACT extraction rate, FALLING falling number, GRAIN, 1000 grain weight, LOAF_H loaf height, LOAF_V loaf volume, NITROGEN N content per grain, PWA water availability, PROT protein content per grain, SDS SDS sedementation, SULFUR S content per grain, TOTAL_N N content per m^2, TOTAL_S S content per m^2, YIELD grain yield per m^2.

between two and eight single nucleotide polymorphisms (SNPs), depending on the probeset.

Discussion

We have demonstrated a novel methodology for the identification of transcripts which show correlation with yield and quality traits in selected lines from a mapping population. As we were interested in traits that were stable across widely varying environments, a full expression QTL study of a whole population would have required hundreds of arrays and huge resources. Furthermore, confounding effects of segregation for development and known alleles with major effects in the population would make analysis of such a study problematic. We therefore used an alternative approach in which selected lines that showed considerable variation in traits were compared at two stages of caryopsis development and were grown in multiple environments. Furthermore, these lines were matched for known determinants of quality (HMW subunits and the 1BL/1RS translocation), allowing novel quality-related transcripts and loci to be identified. The use of multiple environments and years should also ensure that the transcripts identified are robust and not affected by changes in environment. This approach identified almost 470 transcripts which were correlated with genotypic variation in traits, 50 of which were selected to be of greatest relevance and 32 of which were mapped.

If all, or most, of the variation in a trait between the selected genotypes is caused by a single locus controlling expression, we might expect the transcript of the gene to be identified as a *cis* factor in the analysis presented here. In contrast, a trait determined by multiple loci, but associated with the expression of a single gene, would be identified as a transcript controlled by a *trans* factor. In the first case, the sequence encoding the transcript will be close to the sequence causing the trait. In the second case, it is not possible to identify the causative sequence without further analysis.

Validation of the candidate transcripts identified here could be achieved by the generation of near-isogenic lines for one of the traits. If the transcript expression levels continue to segregate with the trait, the expression, or the transcript itself in the case of *cis* factors, could be a valuable marker for the trait. This would justify further investigation to determine whether variation in expression is the functional cause of the trait. Such a finding would also represent a major step forwards in our understanding of genetic determinants of yield and quality, and could allow the development of improved varieties in all backgrounds.

Experimental procedures

Lines

A DH population from the wheat cross Spark × Rialto was used for all mapping and transcriptome studies. Spark is a hard milling, strong gluten, Group 1 breadmaking variety from Nickersons Seeds UK Ltd. (Rothwell, Market Rasen, Lincolnshire, UK), and Rialto is a hard milling, strong gluten, Group 2 variety containing the 1BL/1RS translocation from RAGT Seeds (Saffron Walden, Cambridgeshire, UK). A population of 144 DH lines was developed using the maize cross technique (Snape *et al.*, 2007). Using initial marker characterization, 60 candidate DH lines of the population, all lacking the 1BL/1RS translocation, were taken from a 2001 field experiment at the John Innes Centre (Norwich, UK) and baking tests were performed at RHM Technology (High Wycombe, Buckinghamshire, UK). NIR spectroscopy was carried out for the 60 candidate DH lines using a Perten DA7000 NIR spectrometer (Calibre Control Ltd., Warrington, UK). For each measurement, 20 spectra were collected using a combination of 10 repacks with two repeats each. The parental lines were measured four times at intervals evenly spaced between measurements of the progeny, and all samples from harvests from both years were measured on three separate occasions. The mean spectra were calculated and the spectral range was truncated to 450–1700 nm. The four mean measurements for each parental line were used as a training set to calculate the overall mean spectra, the first principal component (PC) loadings and scores, and the within-group covariance matrix. The mean spectra for the progeny were treated as the test set. They were centred to the mean of the training set, and the PC scores associated with the training set PC loadings were calculated. The Mahalanobis distance from each parent was then calculated using the first PC scores and the inverse of the within-group covariance matrix. This gave an indication of the relative closeness of the offspring to the parents. These data, together with statistical analysis of the test baking data, were subjected to QTL analysis using genotype mapping data developed in previous projects. This identified lines with complementary patterns of quality QTL alleles inherited from each parent.

Plant growth

Seeds of the above lines were sown at the Plant Environment Laboratory, University of Reading on 9–10 December 2003 and 2004 in 18-cm-diameter pots containing 2 : 1 : 2 : 0.5 of vermiculite : sand : gravel : compost mixed with Osmocote slow-release granules (2 kg/m^3) containing a ratio of 15 : 11 : 13 : 2 of N : P_2O_5 : K_2O : MgO. The total fresh weight of the growing medium at sowing was 2.8 kg/pot. Nine seeds per pot were sown, and later thinned to four seedlings per pot. The pots were arranged within two polyethylene-covered tunnels (polytunnels) in a randomized row–column design comprising four replicate 'blocks', each containing 10 (2003) or eight (2004) genotype 'plots', and surrounded by spare pots that were not sampled. Plots comprised 48 and 60 pots in 2003 and 2004, respectively.

From sowing to 14 days post-anthesis (dpa), the polytunnels were maintained close to the outside temperature using fan-assisted ambient air ventilation (with frost protection) with a natural photoperiod and full drip irrigation. Mildewicide and aphicide were applied as and when required, which was no more than twice in any 1 year. At 14 dpa in each year, the environment in one polytunnel

was continued at ambient temperature and full irrigation, whilst, in the second polytunnel, the temperature was controlled at 5–10 °C above ambient (but below 28 °C) and the soil moisture content was allowed to fall to 15%, close to the wilting point. A summary of the temperature and relative humidity measurements during these periods is provided in Table S1 (see Supporting Information).

Plants were tagged at anthesis and ears were harvested at weekly intervals until maturity. Grains from two of these ears were counted before drying at 80 °C for 48 h and determining the mean grain dry weight and the nitrogen and sulphur contents. Ears were also sampled for transcriptome analysis (see below). At maturity, the remaining ears were harvested and threshed. Grain yields were adjusted to an 85% dry matter basis.

Grain filling

Ordinary logistic models (constant omitted) were fitted to genotype means within a tunnel for quantity of dry matter, nitrogen and sulphur per grain. These curves were used to deduce the maximum rate of filling, the fitted value on the final assessment date and the duration of filling, i.e. the time taken to reach 95% of the final fitted value (Gooding *et al*., 2005). The fitted values were then subjected to analyses of variance (ANOVAs) in which the treatment structure was G × E and the block structure was year × tunnel × genotype.

Grain testing

Full details of the small-scale tests used to measure grain quality are available elsewhere (Gooding and Davies, 1997). In summary, thousand grain weights (TGWs) were measured using an automated seed counter with a known weight of grain, dried to the above standard. Specific weight (SWT; bulk packing density) was determined on fresh grain with a chondrometer and screenings by passing 100 g of grain over a 2.0-mm-wide grain sieve for 30 s on a sieve shaker, the weight passing through the sieve being expressed as a percentage of the original sample weight. Samples were milled using a Laboratory Mill 3100 (Perten Instruments AB, Huddinge, Sweden) with a screen of 0.8 mm in diameter. The nitrogen concentration was determined with an oxidative combustion method using an automated Dumas-type analyser (Leco FP-528; Leco Instruments (UK) Ltd., Stockport, Cheshire, UK), and the protein concentration was calculated as N × 5.7. Grain sulphur content was also determined after oxidative combustion with a Leco SC-144DR. The Hagberg falling number (HFN Perten Instruments AB, Huddinge, Sweden) was measured with an apparatus incorporating automatic agitation (HFN apparatus, Stockholm, Sweden). The SDS sedimentation test was performed as an indicator of potential baking performance (BSI ISO/CD 309).

Milling and baking

Full baking tests were conducted by RHM Technology for each genotype × environment × year combination on grain samples bulked over the four blocks in each tunnel. In summary, the flour extraction rate was determined after milling in a Buhler (AG Switzerland, Buhler Ltd., London, UK) test mill with a target extraction rate of 75%, water absorption was determined in a Farinograph test, dough rheology was assessed with a Kieffer (Kieffer, Stable Micro Systems Ltd., Godalming, UK) rig, and loaf volume and texture were assessed after baking 400-g loaves.

Transcriptomic analysis

Ten developing caryopses from the middle part of each ear were harvested at 14 and 23 dpa and immediately frozen in liquid nitrogen before storage at –70 °C for RNA extraction. Each sample comprised 100 caryopses from 10 main stems or first tillers, with single samples of each year × line × stage × environment combination being used for transcriptome analysis. This gave a total of 56 Affymetrix datasets.

Methods for RNA isolation are described by Wan *et al*. (2008). Hybridization to the public Affymetrix wheat Genechip® microarray was conducted by Syngenta Inc. (NC, USA), according to the manufacturer's protocol. Data were normalized as described previously (Wan *et al*., 2008) using the gcRMA algorithm and GeneSpring GX 7 package (Agilent Technologies, Santa Clara, CA, USA). Probesets showing significant absolute expression (> 10) under any condition were selected. This set of 33 000 probesets was then subjected to two-way ANOVA for effects of developmental stage × genotype, effectively treating the different years and tunnels as replicates. From this, probesets which showed significant genotype effects at $P < 0.05$ with Benjamini–Hochberg false discovery rate multiple-testing correction were identified. The set of 1905 probesets showing genotype-dependent expression was then used to look for correlation with quality traits. This is the first step in the identification of candidate transcripts, as shown in Figure 5.

Identification of transcripts correlated with quality traits

There were major environmental effects on nearly all of the traits as a result of the treatments imposed in the tunnels and, to a lesser extent, variation between years (Table S2, see Supporting Information). These were largely explained by effects on the duration of the grain filling period, which was shorter under the hot and dry treatment and in 2003–04, which was warmer than 2004–05. These effects on the grain filling duration would not be expected to be greatly reflected in the transcriptome, partly as they would apply largely after the RNA sampling dates used here. The statistical model relating gene expression to quality traits therefore allowed for separate, additive effects of environmental treatment (tunnel) and year on the measured quality parameter.

A filtering of probeset values was performed on the basis of: (i) quality of measurement; (ii) at least a two-fold change in expression; and (iii) significant variety gene expression differences (after controlling by tunnel effects). This was performed for the 14 and 22-dpa probesets separately. For each selected probeset, a linear regression model relating the logarithmically transformed gene expression to each of the breadmaking and harvest traits individually was performed. Significant F ratios from ANOVA at $P < 0.05$ were used to identify related variables.

The process of selecting candidate transcripts involved in determining traits using the above model is the second step in Figure 5. This identified 468 probesets in which expression was significantly correlated ($P < 0.05$, corresponding to a false discovery rate of 52%) with one or more of the 23 harvest and quality traits (listed in Table S12, see Supporting Information). These were then manually prioritized based on the following criteria: putative gene function from sequence similarity to known genes and biologically plausible link to correlated trait; size of the effect on gene expression; strength of correlation; and importance of trait. This resulted in a set of 50

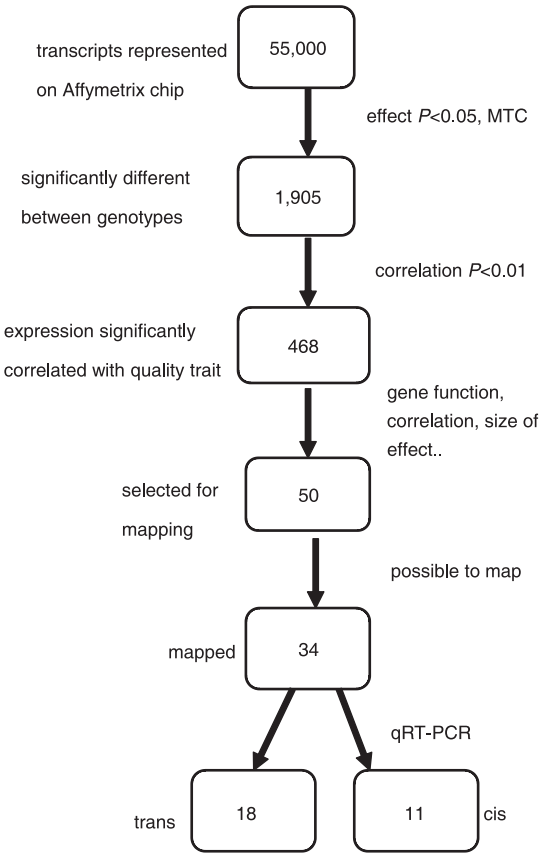

Figure 5 Flow diagram for the identification of candidate transcripts.

transcripts selected as candidates for which allelic differences cause the genotypic variation in the correlated trait.

Transcript mapping

The mapping of transcripts was based on single-stranded conformation polymorphism gel technology, and is the subject of a separate publication (M. Leverington *et al*., unpubl. data). Of the 50 selected transcripts, 34 were mapped within the Spark × Rialto population.

qRT-PCR

Variation in expression between lines was determined for 15 of the mapped transcripts which showed the most promise as candidates by qRT-PCR. The methodology was as described by Wan *et al*. (2008).

Acknowledgements

Rothamsted Research, Institute of Food Research and John Innes Centre receive grant-aided support from the Biotechnology and Biological Sciences Research Council (BBSRC) of the UK. This work was supported under the BBSRC Exploiting Genomics Programme: EGA17694 to Rothamsted Research (Harpenden, Hertfordshire, UK), EGA17705 to John Innes Centre (Norwich, UK), EGA17703 to University of Reading (Reading, Berkshire, UK), EGA17706 to University of Bristol (Bristol, UK) and 218/EGA17713 to Institute of Food Research (Colney, Norwich, UK).

References

Blumenthal, C.S., Barlow, E.W.R. and Wrigley, C.W. (1993) Growth environment and wheat quality: the effect of heat stress on dough properties and gluten proteins. *J. Cereal Sci*. **18**, 3–21.

Cornish, G.B., Békés, F., Eagles, H.A. and Payne, P.I. (2006) Prediction of dough properties for bread wheats. In: *Gliadin and Glutenin: the Unique Balance of Wheat Quality* (Wrigley, C.W., Békés, F. and Bushuk, W., eds), pp. 243–280. St Paul, MN: AACC.

Dupont, F.M. and Altenbach, S.B. (2003) Molecular and biochemical impacts of environmental factors on wheat grain development and protein synthesis. *J. Cereal Sci*. **38**, 133–146.

Gooding, M.J. and Davies (1997) *Wheat Production and Utilisation: Systems, Quality and the Environment*. Wallingford, UK: CABI.

Gooding, M.J., Gregory, P.J., Ford, K.E. and Pepler, S. (2005) Fungicide and cultivar affect post-anthesis patterns of nitrogen uptake, remobilization and utilization efficiency in wheat. *J. Agric. Sci*. **143**, 503–518.

Graybosch, R.A. (2001) Uneasy unions: quality effects of rye chromatic transfers to wheat. *J. Cereal Sci*. **33**, 3–16.

Payne, P.I. (1987) Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Ann. Rev. Plant Physiol*. **38**, 141–153.

Porter, J.R. and Semenov, M.A. (2005) Crop responses to climatic variation. *Philos. Trans. R. Soc. B*, **360**, 2021–2036.

Richter, G.M. and Semenov, M.A. (2005) Modelling impacts of climate change on wheat yields in England and Wales: assessing drought risks. *Agric. Systems*, **84**, 77–97.

Shewry, P.R., Halford, N.G. and Tatham, A.S. (2003) The high molecular weight subunits of wheat glutenin and their role in determining wheat processing properties. In: *Advances in Food and Nutrition Research*, Vol. 45 (Taylor, S.L., ed.), pp. 219–302. London, UK: Academic Press.

Snape, J.W., Foulkes, M.J., Simmonds, J., Leverington, M., Fish, L.J., Wang, Y. and Ciavarella, M. (2007) Dissecting gene x environmental effects on wheat yields via QTL and physiological analysis. *Euphytica*. **154**, 401–408.

Wan, Y., Poole, R.L., Huttly, A.K., Underwood, C., Feeney, K., Welham, S., Gooding, M.J., Mills, E.N.C., Edwards, K.J., Shewry, P.R. and Mitchell, R.A.C. (2008) Transcriptome analysis of grain development in hexaploid wheat. *BMC Genomics*, **9**, Article 121.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Temperature and relative humidity in the two treatments (cool, wet and hot, dry) in 2003–04 and 2004–05 for the entire crop growing period and from when the treatments were imposed at 14 days post-anthesis (dpa)

Table S2 The effect of genotype and environment on the grain yields of dry matter, nitrogen and sulphur

Table S3 Effect of genotype and environment on the thousand grain weight, screenings, specific weight and grain Hagberg falling number of winter wheat

Table S4 Effect of genotype and environment on grain filling with dry matter and nitrogen and sulphur in wheat, as fitted over time by ordinary logistic models

Table S5 Effect of genotype and environment on the grain concentrations of nitrogen and sulphur, and the sodium dodecylsulphate sedimentation volume of winter wheat

Table S6 Effect of genotype and environment on the flour characteristics of winter wheat

Table S7 Effect of genotype and environment on the colour of winter wheat

Table S8 Effect of genotype and environment on the Farinograph test of winter wheat

Table S9 Effect of genotype and environment on dough data for baking

Table S10 Effect of genotype and environment on dough measurements from Kieffer rig

Table S11 Effect of genotype and environment on dough data for baking

Table S12 Harvest and breadmaking varieties analysed for correlation with transcript abundance. *Corrected for temperature variation

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

