Theor Appl Genet (2008) 117:401–411
DOI 10.1007/s00122-008-0784-5

ORIGINAL PAPER

Surveying expression level polymorphism and single-feature polymorphism in near-isogenic wheat lines differing for the *Yr5* stripe rust resistance locus

Tristan E. Coram · Matthew L. Settles · Meinan Wang · Xianming Chen

Received: 29 November 2007 / Accepted: 26 April 2008 / Published online: 10 May 2008

Abstract DNA polymorphisms are valuable for several applications including genotyping, molecular mapping and marker-assisted selection. The 55 K Affymetrix Wheat GeneChip was used to survey expression level polymorphisms (ELPs) and single-feature polymorphisms (SFPs) between two near-isogenic wheat genotypes (BC_7:F_4) that differ for the *Yr5* stripe rust resistance locus, with the objective of developing genetic markers linked to *Yr5*. Ninety-one probe sets showing ELPs and 118 SFP-containing probe sets were identified between isolines, of which just nine ELP probe sets also contained SFPs. The proportion of the transcriptome estimated to be variable between isolines from this analysis was 0.30% for the ELPs and 0.39% for the SFPs, which was highly similar to the theoretical genome difference between isolines of ~0.39%. Using wheat-rice synteny, both ELPs and SFPs mainly clustered on long arms of rice chromosomes four and seven, which are syntenous to wheat chromosomes 2L (*Yr5* locus) and 2S, respectively. The strong physical correlation between the two types of polymorphism indicated that the ELPs may be regulated by *cis*-acting DNA polymorphisms. Twenty SFPs homologous to rice 4L were used to develop additional genetic markers for *Yr5*. Physical mapping of the probe sets containing SFPs to wheat chromosomes identified nine on the target chromosome 2BL, thus wheat-rice synteny greatly enhanced the selection of SFPs that were located on the desired wheat chromosome. Of these nine, four were converted into polymorphic cleaved amplified polymorphic sequence (CAPS) markers between *Yr5* and *yr5* isolines, and one was mapped within 5.3 cM of the *Yr5* locus. This study represents the first array-based polymorphism survey in near-isogenic genotypes, and the results are applied to an agriculturally important trait.

Communicated by M. Sorrells.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0784-5) contains supplementary material, which is available to authorized users.

T. E. Coram (✉) · X. Chen
US Department of Agriculture, Agricultural Research Service, Wheat Genetics, Quality, Physiology and Disease Research Unit, Pullman, WA 99163, USA
e-mail: tristan.coram@ars.usda.gov

T. E. Coram · M. Wang · X. Chen
Department of Plant Pathology, Washington State University, Pullman, WA 99164-6430, USA

M. L. Settles
Department of Molecular Biosciences, Washington State University, Pullman, WA 99164-6430, USA

Introduction

For most plants, DNA polymorphisms are responsible for a significant proportion of phenotypic variation. Subsequently, DNA polymorphisms are valuable for several applications including genotyping, molecular mapping and marker-assisted selection. Associating a particular trait with a DNA polymorphism is especially valuable for breeding purposes. Recently, novel means of identifying abundant DNA polymorphisms have been developed that incorporate a microarray platform for high-throughput genotyping. These include diversity array technology (DArT™) (Jaccoud et al. 2001) and subtracted diversity array (Jayasinghe et al. 2007) for DNA:DNA hybridization-based genotyping and mapping studies. More recently, expression level polymorphisms (ELPs) (Doerge 2002) and single-feature polymorphisms (SFPs) (Rostoks et al. 2005) have been identified from cRNA hybridization studies. In plants, ELPs have been

studied in mapping populations (Kliebenstein et al. 2006a, b; Van Leeuwen et al. 2007; West et al. 2007), and have been associated with various phenotypic traits, for example, seed development in wheat (Jordan et al. 2007), qualitative pathogen resistance in *Arabidopsis thaliana* (Gassmann et al. 1999; Grant et al. 1995) and domestication in maize (Wang et al. 1999). ELPs are known to be caused by both *cis*-acting (Cong et al. 2002; Lambrix et al. 2001) and *trans*-acting factor DNA sequence polymorphisms (Caicedo et al. 2004), as well as expression-altering deletions within a gene itself (Kliebenstein et al. 2001; Lambrix et al. 2001). ELPs between genotypes have been shown to correlate with underlying DNA sequence polymorphism in *Arabidopsis thaliana* (Kliebenstein et al. 2006a) at both a whole-genome and sub-chromosomal level, which indicated a significant impact of *cis*-, or linked *trans*-acting, DNA polymorphism on gene expression.

Single-feature polymorphisms (SFPs) are detected by high-density oligonucleotide probe arrays, and represent DNA sequence polymorphism between genotypes within an individual oligonucleotide probe that can be detected by hybridization affinity difference (Borevitz et al. 2003). Importantly, SFPs occur within a transcribed gene that may reflect variation in phenotype. Recent studies have been successful in identifying DNA polymorphisms from RNA hybridized to Affymetrix GeneChips (Affymetyrix, Santa Clara, CA) in addition to transcript abundance polymorphisms (Borevitz et al. 2003; Cui et al. 2005; Rostoks et al. 2005; Walia et al. 2007; West et al. 2006). Use of RNA acts to reduce genome complexity, which overcomes the difficulties associated with hybridizing genomic DNA from organisms with large and complex genomes such as wheat (*Triticum aestivum* L.). SFPs are often caused by single nucleotide polymorphisms (SNPs), as demonstrated by studies that have sequenced SFPs (Borevitz et al. 2003; Rostoks et al. 2005). However, SFPs are only predictive of a potential SNP within an oligonucleotide probe, so further work must be performed to actually identify the causative SNP. Subsequently, it is now possible to accurately identify both ELPs and SFPs from RNA expression studies using short oligonucleotide arrays. Both SFPs and ELPs can also be mapped as alleles in segregating populations ("genetical genomics"), and ELPs can be considered as e-traits to establish expression QTL (eQTL) (Jordan et al. 2007; West et al. 2006, 2007). In wheat, one recent eQTL study has been reported, which utilized the Affymetrix Wheat GeneChip to map ELPs for seed development in a population of 41 doubled haploid lines, but did not map SFPs (Jordan et al. 2007). By using wheat-rice synteny, the authors found that most major effect eQTL were *cis*-acting, which has also been observed in eQTL studies of other organisms, including *A. thaliana* (West et al. 2007) and the mouse (Schadt et al. 2003).

The objective of this study was to determine the level of ELP and SFP in two near-isogenic wheat genotypes (BC_7:F_4) that differ for stripe rust (*Puccinia striiformis* Westend. f. sp. *tritici* Eriks. (*Pst*)) resistance controlled by a single gene known as *Yr5*. The *Yr5* gene was originally identified in *T. spelta* var. *album* (TSA) (Chen et al. 1998; Lupton and Macer 1962; McIntosh et al. 1998), and confers race-specific all-stage resistance to all currently known races of *Pst* in the United States (Chen, unpublished data). The chromosomal location of *Yr5* has also been determined on the long arm of chromosome 2B (Law 1976; Macer 1966). For this study, the aim was to identify ELPs and SFPs from an Affymetrix Wheat GeneChip dataset used as part of a previous transcript abundance study of the response of each isoline to *Pst* inoculation (Coram et al. 2008). The goal was to assess the level of ELP and SFP in the near-isogenic genotypes across two treatments (*Pst*-inoculation and mock-inoculation), and to discover polymorphisms linked to the *Yr5* locus. This study represents the first of its kind to be applied to near-isogenic genotypes, as well as the first survey of SFP in wheat. Importantly, the results demonstrate the utility of SFPs for generating genetic markers linked to an agriculturally important trait.

Materials and methods

Wheat GeneChip

The GeneChip® Wheat Genome Array (Affymetrix, Santa Clara, CA) is a 3′IVT array that includes 61,127 probe sets representing 55,052 transcripts for all 21 wheat chromosomes in the genome. 59,356 probe sets represent modern hexaploid (A, B and D genomes) wheat (*T. aestivum*) and are derived from the public content of the *T. aestivum* UniGene Build #38 (April 24, 2004). 1,215 probe sets are derived from ESTs of a diploid near relative of the A genome (*T. monococcum*), a further 539 represent ESTs of the tetraploid (A and B genomes) durum wheat species *T. turgidum*, and five are from ESTs of a diploid near relative of the D genome known as *Aegilops tauschii*. Probe sets consisted of pairs of 11 perfect match (PM) and mismatch (MM) 25-mer oligonucleotides designed from the 3′ end of exemplar sequences, with nucleotide 13 as the MM. Each probe set was assumed to represent a transcript and array annotation information is available on the NetAffx data analysis center (http://www.affymetrix.com).

Plant material and GeneChip data

Near isogenic lines (NILs) for the *Yr5* resistance gene were developed at the Plant Breeding Institute, Sydney, Australia, by backcrossing the *Yr5* donor [*T. aestivum* subsp.

spelta (L.) Thell. cv. Album (TSA)], with the recurrent susceptible spring wheat genotype (*T. aestivum* L.) "Avocet Susceptible" (AVS) seven times [BC_7:F_4 AVS × (AVS × TSA)] and selecting for the appropriate resistance in each generation (Wellings et al. 2004). Backcrosses were advanced to the BC_7:F_4 stage (Yan et al. 2003), thus only ~0.39% of the TSA genome was remaining in the *Yr5* NIL in the AVS background. Two BC_7:F_4 NILs were selected that differed at the *Yr5* locus; *Yr5* (homozygous resistant) and y*r5* (homozygous susceptible). As part of a previous gene expression study (Coram et al. 2008), GeneChip data was generated from RNA of 10-day-old leaf tissue of the two isolines in an experimental design that included two treatments (mock-inoculation and *Pst*-inoculation), four time-points (6, 12, 24, and 48 h) and three biological replications. Subsequently, 24 total GeneChip hybridizations were made for each isoline according to standard Affymetrix protocols. The transcript data was validated by quantitative RT-PCR in the previous study (Coram et al. 2008). All minimum information about microarray experiments (MIAME) guidelines were observed (Brazma et al. 2001) and GeneChip data can be found in WheatPLEX (Shen et al. 2005), accession number TA9.

Expression level polymorphism analysis

The Bioconductor (Gentleman et al. 2004) package *affy* was used to read in the 48 raw Affymetrix ".CEL" files, and quality control was assessed using the *affyQCreport* (Parman C, Halling C, Gentleman R affyQCReport:QC Report Generation for affyBatch objects. R package version 1.16.0.) of Bioconductor (Gentleman et al. 2004), which provided Affymetrix recommended quality metrics, per array intensity distributions, between array comparisons, and other diagnostic plots. The data was pre-processed using robust multi-array average (RMA) (Bolstad et al. 2003; Irizarry et al. 2003), and hierarchical clustering (Euclidean metrics, complete linkage) of normalized arrays using the *hclust* function in R (R Development Core Team 2006) revealed that the RMA algorithm accurately separated the variables of the data set (genotype, treatment, time and replicate). The data set was then separated into time-points and a linear model (Smyth 2005) was applied on each probe set to detect significantly ($P < 0.05$) different transcript levels (ELPs) for isoline, treatment, and the isoline × treatment interaction. Probability (P) values were adjusted for multiple comparisons using false discovery rate (FDR) $\alpha < 0.05$ (Benjamini and Hochberg 1995). Putative ELP probe sets between *Yr5* and *yr5* significant for treatment were identified and annotated using HarvEST (Affymetrix Wheat1 Chip version 1.52), which identified the corresponding unigene for each probe set and provided the current best BLASTX hit from the non-redundant (nr) database of NCBI, as well as the best BLASTX hits from rice and *A. thaliana* TIGR databases (http://www.tigr.org/plantProjects.shtml). A database hit <1e-10 was considered as significant, otherwise the unigene was annotated as unknown. For gene ontology (GO), the rice locus matching each probe set in the HarvEST output was queried using the TIGR rice genome annotation (Yuan et al. 2003), which provided GO terms including biological function. Unigenes were assigned to functional categories based on Munich Information Center for Protein Sequences (MIPS; http://mips.gsf.de/) classifications.

Expression level polymorphisms were differentiated as qualitative (caused by hybridization in only one isoline) or quantitative (caused by significantly different transcript abundance between isolines) based upon corresponding MAS 5.0 PMA (present, marginal and absent) calls generated by GCOS v.1.4 (Affymetrix, Santa Clara, CA). Probe sets were only identified as being expressed if they were called as "present" in all three biological replications for a given time-point, otherwise they were identified as not expressed. Qualitative probe sets for mock- and *Pst*-inoculation were then identified as those that were expressed in at least three time-points for one isoline and not expressed in all time-points for the other isoline. Additionally, the PMA calls were used to detect the total number of expressed probe sets for which successful hybridization occurred amongst any isoline and condition.

Single-feature polymorphism prediction

Sequence polymorphisms within individual oligonucleotide probes were detected as single-feature polymorphisms (SFPs). To detect SFPs, only the eleven perfect match probes from each probe set were analyzed. The Bioconductor package *affy* was used to read in the 48 raw Affymetrix ".CEL" files from both isolines, and a novel method for SFP detection was applied (code is available from corresponding author upon request). Briefly, the raw probe level signal data were firstly RMA background corrected and normalized with the Bioconductor *affy* package. SFP data was calculated by subtracting the RMA pre-processed expression estimate for a particular probe set from the corresponding probe level signal data. Residuals of the model, which had been normalized for probe and expression effects, were fitted for an isoline effect using the Bioconductor package *siggenes*, where significance analysis of microarrays (SAM) (Tusher et al. 2001) with an FDR adjustment of $\alpha < 0.01$ was used to detect significantly different features. SAM output included *d*-statistics, which represents a modified *t*-test, that were used to assess significance of SFPs. Probe sets containing SFP probes were annotated as described for the ELPs.

Physical and genetic mapping

Using HarvEST annotation, ELPs and SFPs that were homologous to physical map loci in rice were mapped to rice chromosomes. Additionally, ELPs and SFPs that were homologous to wheat deletion bin mapped ESTs from the NSF EST deletion mapping project (Qi et al. 2004) were identified using GrainGenes-SQL (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi). An additional 20 SFPs that were homologous to sequences on rice chromosome 4L (syntenous to 2L in wheat; *Yr5* locus) were assigned to wheat chromosomes using the set of wheat nulli-tetrasomic Chinese Spring lines (Wheat Genetic and Genomic Resources Center, Kansas State University, KA). PCR primers for selected probe sets containing SFPs were designed on corresponding unigenes identified in HarvEST using Vector NTI (v. 10.3.0, Invitrogen Corporation). Each 50 µL PCR reaction consisted of; 60 ng DNA, 0.4 µM forward and reverse unigene-specific primer, 1× GoTaq Flexi buffer (Promega Corporation), 0.2 mM each dNTP, 3.0 mM $MgCl_2$, and 1.5 units GoTaq DNA Polymerase (Promega Corporation). Thermocycling parameters were optimized at 95°C for 5 min, followed by 40 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 1 min 30 s, with a final extension at 72°C for 10 min. PCR products were analyzed by gel electrophoresis for successful amplification and chromosome locations were determined by the amplification pattern of the desired product amongst the nulli-tetrasomic Chinese Spring samples. To enrich for polymorphism, PCR products were separately treated with five different restriction enzymes that have 4-bp recognition sequences; *Hae*III, *Hpa*II, *Alu*I, *Dpn*I and *Rsa*I. Each 20 µL digestion reaction consisted of; 10 µL PCR product, 1× NEB buffer (New England BioLabs, Ipswich MA), 10 units restriction enzyme (New England BioLabs, Ipswich MA), and 100 µg/mL BSA if required (New England BioLabs, Ipswich MA). Each digestion was incubated at 37°C for 2 h and then inactivated at 80°C for 20 min. Products were visualized by 2.8% high-resolution agarose gel electrophoresis and chromosome locations were determined as described above.

Using the same PCR and restriction digestion conditions, the 20 selected SFP primer sets were subsequently used for amplification on *Yr5* and *yr5* genomic DNA, as well as resistant and susceptible DNA bulks from a BC_7:F_2 population generated from the BC_7 AVS × (AVS × TSA) cross (M. Wang, unpublished). The DNA bulks were randomly derived from 20 resistant and 20 susceptible lines. Digested PCR products were visualized using 2.8% high-resolution agarose gel electrophoresis to detect potential polymorphisms. Polymorphic cleaved amplified polymorphic sequence (CAPS) markers between *Yr5* and *yr5* DNA, as well as resistant and susceptible DNA bulks, were screened on 114 BC_7:F_3 lines developed from the BC_7 AVS × (AVS × TSA) cross that was previously used to develop *Yr5* STS markers (Chen et al. 2003).

Results

Expression level polymorphisms

The linear model analysis with false discovery rate (FDR) multiple-testing correction (Benjamini and Hochberg 1995) detected significant gene expression differences (ELPs) between isolines for the mock-inoculation and *Pst*-inoculation treatments. We found 77 ELPs between isolines for mock-inoculation and 81 ELPs for *Pst*-inoculation (Supplementary Table 1, in Electronic Supplementary Material). Sixty-seven of the ELPs were detected under both treatments, whereas 10 were specific to mock-inoculation and 14 to *Pst*-inoculation. This result suggested that *Pst*-inoculation did not strongly influence expression of the leaf transcriptome as a whole. Approximately half the total ELPs (45) had significantly higher expression in the *Yr5* isoline compared to *yr5*. Of these, 34 (76%) were due to quantitative transcript differences and 11 (24%) were considered as qualitative (see "Materials and Methods"). For the 46 ELPs expressed more highly in *yr5*, 26 (57%) were quantitative and 20 (43%) were qualitative. HarvEST annotation of ELP probe sets revealed that 50 (55%) possessed a significant match (<1e-10) to a hypothetical or functional protein. For the 67 ELPs common to both treatments, 35 could be annotated, of which a high proportion were related to plant defense and signal transduction (40%). Nine of the 14 *Pst*-specific ELPs were annotated, with most of these (66%) related to defense or signal transduction, as expected. The annotations of the mock-inoculation specific transcripts did not show any trend toward a certain functional category. When the data was separated by isoline, the annotations were dominated by defense and signal transduction (39% for *Yr5* and 48% for *yr5*). To measure the proportion of the transcriptome identified as ELPs, we calculated the number of probe sets on the GeneChip that were considered as expressed in at least one experimental condition by again using the MAS 5.0 PMA calls (see "Materials and Methods"). Across all hybridizations, 30,102 out of the 61,127 GeneChip probe sets were "present" in at least one condition. This translated into 0.30% of the detected transcriptome considered as ELPs, which is in line with the ~0.39% theoretical genome difference between the isolines.

Single-feature polymorphisms

A custom script was written to detect SFPs between isolines (code is available from corresponding author upon request), and identified 297 SFPs at a <0.1% false discovery

Table 1 The 20 single-feature polymorphisms (SFPs) between *Yr5* and *yr5* wheat isolines targeted for genetic marker development, where "Deletion bin" refers to wheat ESTs mapped by the NSF EST deletion mapping project (Qi et al. 2004), "Wheat chromosome" indicates the result of the Chinese Spring nulli-tetrasomic analysis with corresponding restriction enzyme used, "Isoline" indicates polymorphism between isoline DNA, and "Bulks" indicates polymorphism between resistant and susceptible bulks from the *Yr5* segregating population

Probe set ID	Deletion bin	Wheat chromosome	Isoline	Bulks
Ta.13376.1	NA	NA	No polymorphism	No polymorphism
Ta.13412.1	NA	NA	NA	NA
Ta.14710.1	NA	2B (*HpaII*)	Yes (*HpaII*)	No polymorphism
Ta.1705.1	NA	NA	No polymorphism	No polymorphism
Ta.21235.1	NA	NA	No polymorphism	No polymorphism
Ta.25199.1	NA	2B (*HaeIII*)	No polymorphism	No polymorphism
Ta.25512.1	NA	2D (*HaeIII*)	No polymorphism	No polymorphism
Ta.25528.1	NA	2B (*HpaII*)	Yes (*HpaII*)	No polymorphism
Ta.28321.2	NA	NA	No polymorphism	No polymorphism
Ta.2884.1	NA	2B (*HaeIII*)	Yes (*HaeIII*)	No polymorphism
Ta.5622.2	NA	2B (*HaeIII*)	NA	NA
Ta.6979.1	2AL 2BL (bin 0.89–1.00) 2DL	2D (*RsaI*)	Yes (*RsaI*)	Yes (*RsaI*)
Ta.8144.1	2AL 2BL (bin 0.50–0.89) 2DL	2B (*HaeIII*)	No polymorphism	No polymorphism
Ta.8151.1	2AL 2BL (bin 0.89–1.00) 2DL	2D (*HpaII*)	No polymorphism	No polymorphism
Ta.9253.1	1AS	2B (*HpaII*)	Yes (*HpaII*)	No polymorphism
Ta.9428.1	2AL 2DL	NA	Yes (*HpaII*)	No polymorphism
Ta.9452.1	NA	2B (*HpaII*)	NA	NA
Ta.9920.1	NA	2B/2D (*HaeIII*)	No polymorphism	No polymorphism
TaAffx.123757.1	2AL 2BL (bin 0.36–0.50) 2DL	2B (*RsaI*)	No polymorphism	No polymorphism
TaAffx.52427.1	NA	NA	NA	NA

NA indicates the absence of available data

rate. Figure 1 shows an example of a normalized expression and residual plot fitted for isoline differences for a probe set in which a SFP was identified at probe 10. Residual plots were also visually inspected to confirm each SFP. The absolute magnitude of the SAM (Tusher et al. 2001) *d*-statistic was used as a threshold to call significant SFPs, and the sign indicated which isoline was polymorphic with regard to the reference GeneChip oligo (positive values predicted SFPs in *Yr5* and negative values in *yr5*). Figure 2 shows the distribution of *d*-statistics for all probes, where significant SFPs are indicated as probes above and below the cutoff lines that represent FDR $\alpha < 0.01$. Additionally, Fig. 2 shows a truncated histogram of the *d*-statistic distribution for all probes, where *Yr5* SFPs are identified in the right tail and *yr5* SFPs in the left tail. Complete results of the SFP analysis for each significant probe, including *d*-statistics and *q* values can be found in Supplementary Table 2 (ESM). The 297 individual SFP probes were sourced from 118 probe sets, where most probe sets contained only one SFP probe (58%) and there was one instance of all 11 probes being called SFPs in a single probe set (Supplementary Table 3, in ESM). Approximately half (45%) of the significant SFP probes identified *Yr5* as the polymorphic genotype, whilst 55% identified *yr5* as polymorphic, with respect to the reference probe on the GeneChip. Within probe sets with multiple SFP probes, there were instances of some probes being identified as *Yr5* polymorphic and some as *yr5* polymorphic. Of the 118 probe sets containing SFPs, 99 (84%) possessed a significant match (<1e-10) to a hypothetical or functional protein. Unlike the ELPs, the functional distribution of the probe sets containing SFPs did not show a bias toward defense-related transcripts. As for the ELPs, we estimated the proportion of the transcriptome identified as SFPs as 0.39%, which is also highly sim-

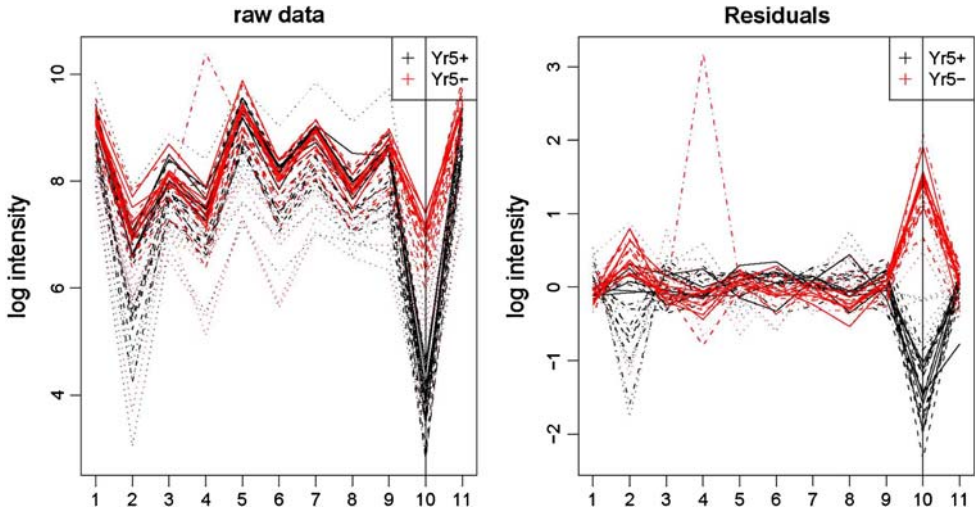

Fig. 1 Normalization of hybridization intensity for single-feature polymorphism (SFP) determination between *Yr5* and *yr5* wheat isolines, where the *y*-axis is background-corrected normalized log intensity data and the *x*-axis is the position of the 11 probes along the probe set. *Black lines* indicate the *Yr5* array signal data for each *Yr5* hybridization, and *red lines* are the *yr5* data. The "raw data" plot shows robust multi-array average (RMA) background corrected and normalized signal data. The "Residuals" plot shows signal data after removing expression effects, which was used for SFP determination and indicates the presence of a SFP in probe 10 of this example

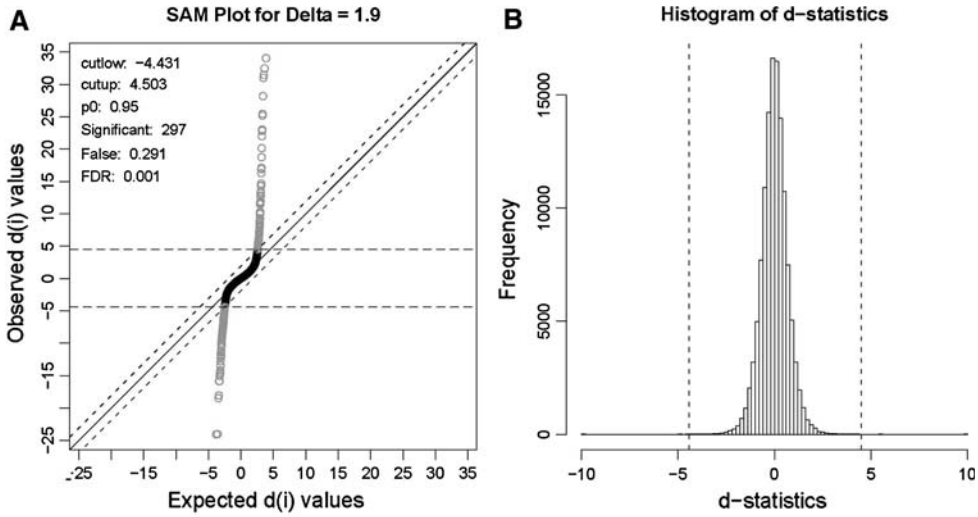

Fig. 2 Distribution of the single-feature polymorphisms (SFPs) detected between the *Yr5* and *yr5* wheat isolines, showing a significance analysis of microarrays (SAM) plot with observed *d*-statistics (*y*-axis) plotted against expected *d*-statistics (*x*-axis). Significant SFPs exceeding the $\alpha < 0.01$ false discovery rate (FDR) are represented by *grey circles*. A histogram of *d*-statistics is also shown, where positive scores >4.503 are *Yr5* SFPs and negative scores <4.431 are *yr5* SFPs

ilar to the ~0.39% theoretical genome difference between the isolines.

Comparing the ELP and SFP lists identified nine probe sets out of 91 (10%) that were called as ELPs and also contained SFPs (Supplementary Table 3 in ESM). To determine if the presence of SFPs influenced the ELP for these nine probe sets, or vice versa, we visually inspected the normalized expression and residual SFP plots (as in Fig. 1) for each probe set (see Supplementary Fig. 1, in ESM). If the expression plot for each isoline across the 11 probes differed by consistent magnitude, we determined this to be a true ELP and as containing false positive SFP(s). The detection of SFPs in these cases was likely due to excessive signal variability across the probes that adversely affected the ability of the algorithm to determine an accurate summarization expression value for the whole probe set. If in the expression plot we saw a few probes with low signal in one isoline only, but the remaining probes had a similar signal value between isolines, we identified these as true SFP(s) for the low signal probes. In these cases the ELP was caused by the low hybridization value of the few probes, which caused a significant gap in the expression summarization values between isolines for the probe set as a whole. In the third observation we saw both a legitimate SFP at the called probe(s) but also saw consistent signal difference amongst the remaining probes. Subsequently, these probe sets were identified as both ELPs and as containing SFPs. After applying these criteria, two of the nine probe sets were classified as true ELPs containing false positive SFPs (Ta.12382.2.S1_at and Ta.26970.1.A1_at),

four represented false positive ELPs and true SFPs (Ta.23313.1.S1_at, Ta.28185.1.S1_x_at, Ta.28363.3.S1_x_at and Ta.28564.1.A1_at), and three were both true ELPs and true SFPs (Ta.30751.1.S1_at, Ta.9452.1.A1_at and TaAffx.54029.1.S1_at).

Physical and genetic mapping

Using HarvEST annotation, it was possible to physically map many ELP probe sets and probe sets containing SFPs in the rice genome (see Supplementary Tables 1 and 3 in ESM). The distribution of ELPs and SFPs amongst rice chromosomes showed a bias toward chromosomes four and seven (Fig. 3). Further, the physical positions on these chromosomes revealed a further bias toward the long arms of these chromosomes (Fig. 4). Rice 4L and 7L are known to be syntenous with wheat chromosome 2L and 2S, respectively (Sorrells et al. 2003). Thus, the isolines used in this study are most polymorphic in the physical region of the *Yr5* locus (2BL). In particular, the SFPs on rice 4L tend to cluster in two distinct regions, the first cluster of six SFPs spans 2.7 Mbp between physical coordinates 19420521-22135874, and the second cluster of 14 SFPs spans 3.6 Mbp between coordinates 30742528–34296287. Also, the ELPs and SFPs tend to co-locate to similar regions in both rice chromosome 4L and 7L.

Probe sets containing SFPs that possessed homology to rice 4L were focused upon for further characterization, and potential genetic marker development, because they were likely to represent sequence polymorphisms (e.g. SNPs) that may be linked to the *Yr5* locus. Of the 20 probe sets containing SFPs that mapped to rice 4L (Table 1), five had been mapped to wheat homeologous group 2L and one to wheat chromosome 1AS by Qi et al. (2004) in the NSF EST deletion mapping project. To further assign each SFP to a particular wheat chromosome, the Chinese Spring wheat nulli-tetrasomic lines were used. PCR primers were designed according to unigene sequences of each probe set containing SFPs and PCR was performed across all nulli-tetrasomic lines. To enrich for polymorphism between PCR products, restriction digestion was performed with a range of restriction nucleases with 4-bp recognition sequences. Thirteen SFPs were successfully assigned to wheat chromosomes, of which nine were placed on 2B, three on 2D, and one on both 2B and 2D (Table 1). Subsequently, using wheat-rice synteny greatly enhanced the selection of SFPs that were located on the desired chromosome in this study. Where available, the data matched that of Qi et al. (2004) in all but one case, where we assigned unigene Ta.9253.1 to 2B instead of 1AS. Of the nine SFPs located on wheat 2B, four were also polymorphic between isolines under the same PCR and digestion conditions. Additionally, one SFP from 2D and one SFP that could not be assigned a chromosome were polymorphic between isolines. These polymorphic SFPs were considered as potential cleaved amplified polymorphic sequence (CAPS) genetic markers, and were screened against resistant and susceptible DNA bulks of the BC_7:F_2 population derived from the BC_7 AVS × (AVS × TSA) cross. CAPS polymorphism was identified in the bulks for Ta.6979.1 (*RsaI*; Chinese Spring 2D), where the pattern of restriction fragments revealed a unique band for *Yr5* that could be mapped as a dominant *Yr5* marker To establish an estimate of genetic linkage to the *Yr5* phenotype, this CAPS marker was screened in a population of 114 lines from a BC_7:F_3 population (Chen et al. 2003), where we detected six recombinant lines that translated into a genetic distance of 5.3 cM between *Yr5* and this marker (data not shown).

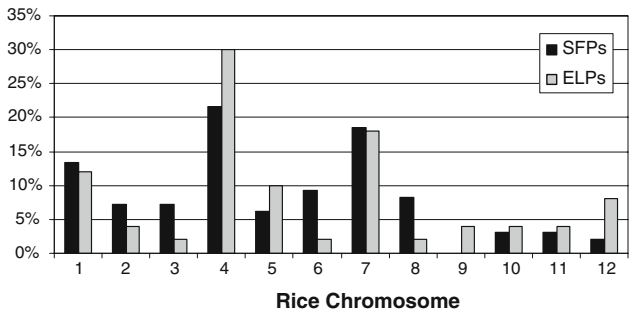

Fig. 3 Histogram showing the proportion of expression level polymorphisms (ELPs) and single-feature polymorphisms (SFPs) detected between the *Yr5* and *yr5* wheat isolines that were homologous to rice physically mapped sequences according to rice chromosome

Discussion

This study identified 91 significant ELPs between two BC_7:F_4 near-isogenic wheat genotypes over two treatments, including *Pst*-inoculation and mock-inoculation, and four sampling time-points. Before identifying ELPs, the RMA condensing algorithm was selected after examining hierarchical clustering of normalized arrays (see "Materials and methods"). However, it is important to recognize that the use of other condensing algorithms (e.g. MAS 5.0) could produce varying lists of ELPs, but for the marker development purpose of this study the RMA algorithm performed satisfactorily. Sixty-seven of the ELPs were conserved between the two treatments, indicating that *Pst*-inoculation did not alter the leaf transcriptome (based on the *Yr5* genotype) to a large extent. However, a large proportion of the ELPs identified for both treatments were related to defense and/or signal transduction pathways, which may be expected considering that the isolines were developed to differ only for their *Pst* response. Other studies surveying ELPs in *A. thaliana* have also found a bias toward genes

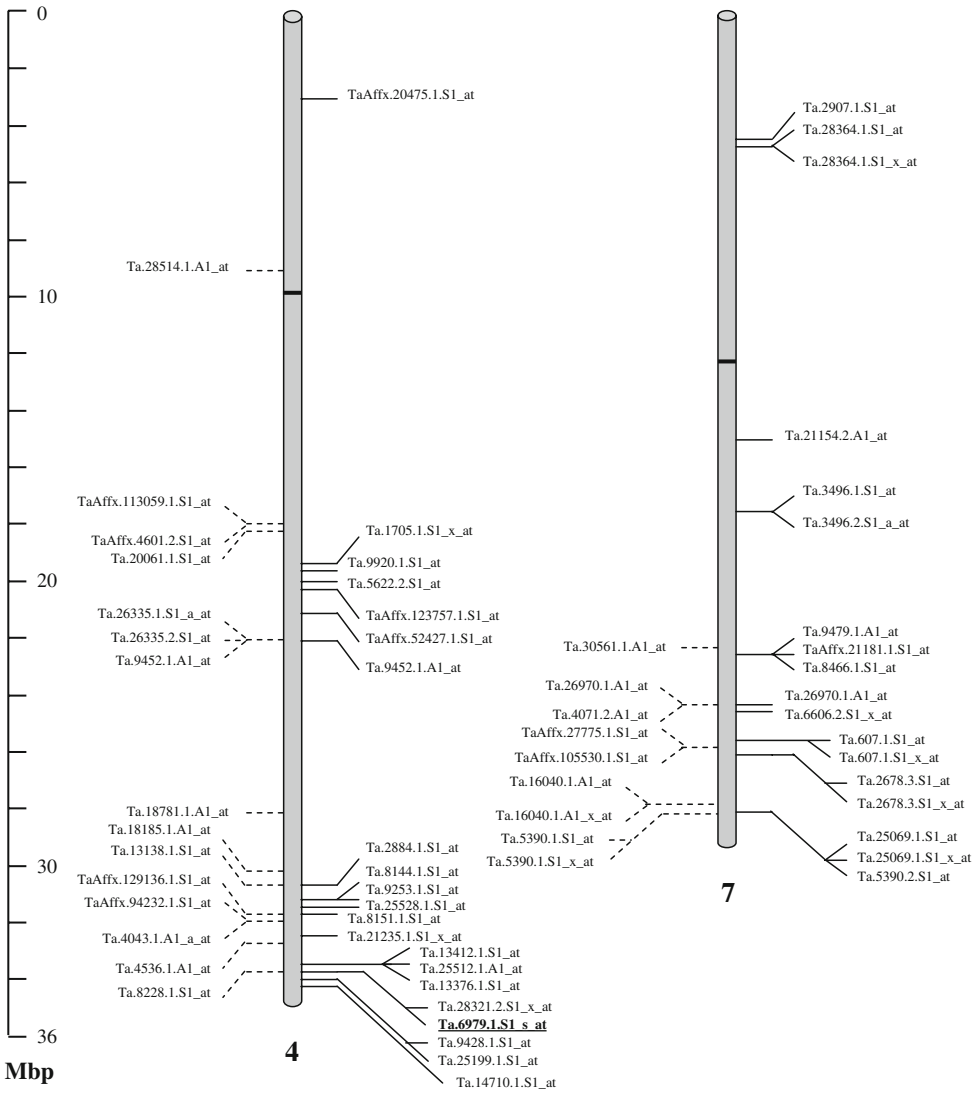

Fig. 4 Physical position of the single-feature polymorphisms (SFPs; *solid lines*) and expression level polymorphisms (ELPs; *broken lines*) detected between the *Yr5* and *yr5* wheat isolines that were homologous to sequences on rice chromosomes 4 and 7. Most ELPs and SFPs cluster on rice 4L and 7L, which are syntenous to wheat chromosomes 2L (*Yr5* locus) and 2S, respectively (Sorrells et al. 2003). SFP Ta.6979.1.S1_at that was mapped within 5.3 cM of the *Yr5* locus is indicated in *bold*

involved in signaling and stress responses (Chen et al. 2005; Kliebenstein et al. 2006a; Van Leeuwen et al. 2007), although salicylic acid was included as a treatment in some. Separating the ELPs into qualitative and quantitative categories for each isoline revealed that *yr5* possessed a larger number of qualitative ELPs in comparison to *Yr5*. We postulated that this observation is most likely due to the fact that the *Yr5* introgression was derived from a different species, *Triticum spelta* var. *album*, thus some of the DNA sequences carried over from this introgression are unlikely to hybridize successfully to the sequences of the Wheat GeneChip, which were sourced from common wheat. Along these lines, it is also likely that *Yr5*, and to a lesser degree *yr5*, possesses additional ELPs that could not be detected by the Wheat GeneChip due to both sequence incompatibility in *Yr5* and the less than whole genome coverage of the Wheat GeneChip. It is also worth noting that the *Yr5* qualitative ELPs are not likely to contain the actual *Yr5* gene due to the derivation of this gene from *Triticum spelta* var. *album*. Expression level polymorphism frequency could be increased by sampling other tissues at different developmental stages under different environmental conditions. However, for the proportion of the transcriptome assessed in this study, we estimated 0.30% as ELPs, which is well supported by the expected genome difference between isolines of ~0.39%.

As another measure of polymorphism between isolines related to actual DNA sequence polymorphism, we utilized a SFP prediction method to identify DNA sequence polymorphism from Affymetrix RNA expression data. Employing this method, we identified 118 probe sets that contained at least one SFP amongst the 11 probes. The detection of 118 SFPs demonstrates the value of this technique for generating abundant polymorphic markers, even between isolines. Thus, SFPs will be valuable in developing dense genetic marker sets for high definition mapping. It is impor-

tant to note that, because of the polyploid nature of the wheat genome, there is potential for cross-hybridization of homeologous gene targets to individual probes, which may affect the specificity of SFP calling. Without a genome sequence this issue is difficult to resolve, but we successfully assigned some probe sets containing SFPs to specific chromosomes using the Chinese Spring wheat nulli-tetrasomic stocks (discussed later). Unlike the ELPs, functional annotation of the probe sets that contained SFPs did not reveal any bias toward stress-related or signaling genes. Because the SFPs were not likely a result of expression differences between isolines, the functional distribution is likely to show more of a random distribution. Just nine (10%) of the ELPs contained SFP(s) and, after visual inspection of the expression and residual plots of these nine probe sets, we found evidence that just four (4%) of the ELPs were likely caused by the SFP(s) in that probe set (false positive ELPs). Although Kliebenstein et al. (2006a) reported that SFPs do not significantly effect ELPs in *A. thaliana*, successfully separating ELPs and SFPs is important for "genetical genomics" studies so as significant expression quantitative trait loci (eQTL) are not confounded by SFP effects, and also so that SFP hybridization-based mapping is not influenced by ELP. The identified SFPs represented 0.39% of the surveyed transcriptome, which was highly similar to the proportion of ELPs and the estimated genome difference between isolines. Because the reference probes on the wheat GeneChip were sourced from different wheat genotypes than those used in this study, cases of DNA polymorphism in the same probe of both isolines could not be detected. Subsequently, as for the ELPs, additional SFPs are also likely to exist, but the correlation to estimated isoline genome difference indicates that the methods of ELP and SFP detection are likely to accurately represent genome differences between genotypes.

Mapping of the ELPs and SFPs to rice indicated that most were located on rice chromosomes 4L and 7L. Rice 4L is syntenous to wheat 2L, whilst rice 7L is syntenous to wheat 2S (Sorrells et al. 2003). This result suggests that wheat homeologous group two was under selection during the development of the isolines, which is expected considering the known position of the *Yr5* locus on wheat 2BL (Law 1976; Macer 1966). In rice, the ELPs and SFPs both mapped to similar regions (Fig. 4), which suggests a positive local correlation between the two types of polymorphism. Other studies of this type have also found evidence for a relationship between DNA sequence divergence and ELPs (Kliebenstein et al. 2006a; Van Leeuwen et al. 2007). Kliebenstein et al. (2006a) suggested that this correlation may be due to a strong influence of *cis*-acting DNA sequence polymorphisms in regulating ELPs. Subsequently, we also conclude that the enrichment for ELPs in the *Yr5* region, where DNA sequence polymorphism is high, indicates that there exists a strong *cis*-, or linked *trans*-, effect of DNA sequence polymorphism on gene expression. Thus, many of the ELPs are likely to harbor DNA polymorphisms in their promoter regions, which could not be assessed by the wheat GeneChip. In particular, the qualitative ELPs resemble the gene expression markers (GEMs) reported by West et al. (2006), which show a distinct bimodal distribution between genotypes and are generally regulated by *cis*-acting DNA polymorphism. The existence of such polymorphisms in the identified ELPs could be determined in future studies by isolating these sequences from the *Yr5* bacterial artificial chromosome (BAC) library (Ling and Chen 2005) and sequencing the promoter regions from both isolines. Alternatively, some quantitative ELPs may result from DNA polymorphisms in linked *trans*-acting factors. Interestingly, twelve (10%) of the probe sets containing SFPs were functionally annotated as transcription associated genes (see Supplementary Table 3 in ESM), which could possibly regulate some of the observed ELPs. Also, some of the 37 unknown or unclear probe sets containing SFPs may represent *trans*-acting factors. However, analysis of SFP influence on ELPs in a segregating population would be required to determine this effect.

Considering that SFPs are known to represent DNA sequence polymorphisms (see review by Zhu and Salmeron 2007), we focused on these as potential genetic markers between *Yr5* and *yr5*, which may be useful for fine genetic mapping and map-based cloning of *Yr5*. The ELPs were not chosen for genetic marker development as they are less likely to possess DNA sequence polymorphism in their transcribed regions. Using the knowledge that *Yr5* is located on wheat 2BL, we chose to further characterize only the 20 SFPs that mapped to rice 4L (syntenous to wheat 2L). Placement of these SFPs onto wheat chromosomes was achieved using Chinese Spring nulli-tetrasomic wheat lines (Wheat Genetic and Genomic Resources Center, Kansas State University, KA), and 9 of the 13 that could be mapped were specific to the targeted chromosome arm, wheat 2BL. Considering the potential for SFPs representing SNPs, a CAPS approach to finding polymorphism was more successful. After screening the nine genetic marker candidates on the parental isolines (*Yr5* and *yr5*) and bulks from a segregating population, we identified six candidates with polymorphism between parents. However, polymorphism in the resistant and susceptible bulks was considered more informative for developing *Yr5* markers, and we identified one marker (Ta.6979.1) that met this requirement. The remaining five were not polymorphic between bulks, thus were not likely to be closely linked to the *Yr5* locus and were not mapped. Using a previously developed population (Yan et al. 2003), we estimated the one marker polymorphic in the bulks (Ta.6979.1) to be

5.3 cM from *Yr5*, which indicates that the assignment of this marker to wheat chromosome 2D was incorrect. Although not as close to *Yr5* as other published markers (Chen et al. 2003; Smith et al. 2007), Ta.6979.1 represents an expressed gene and may be useful for marker-assisted selection, and demonstrates the ability of using SFPs for generating PCR-based genetic markers. Importantly, the marker resides in the 3.6 Mbp cluster of SFPs that were physically mapped to rice chromosome 4L (Fig. 4), which indicates that the ELPs and SFPs are enriched around the *Yr5* locus. Using the TIGR rice genome browser (http://www.tigr.org/tigr-scripts/osa1_web/gbrowse/rice/), we identified seven nucleotide-binding-site/leucine-rich-repeat plant resistance genes and >10 receptor protein kinase genes within this cluster in the rice genome, which we will use in future wheat-rice synteny studies aimed at map-based cloning of the *Yr5* gene.

In summary, this study demonstrates that ELP and SFP surveying and the use of wheat-rice synteny are valuable techniques for generating genetic markers linked to traits of interest in a complex genome. In particular, the use of isolines to survey SFPs enables the development of genetic markers for a trait of interest without profiling an entire population segregating for that trait. At this stage however, the economic constraint of microarray studies means that studies such as this would likely only be performed as part of a larger gene expression study, as was the case here. The results show that the ELPs co-locate to the region of DNA sequence polymorphism between isolines, which suggests that most ELPs are regulated by *cis*-, or linked *trans*-, acting DNA sequence polymorphisms. Further, the SFP-containing probe sets involved in transcription regulation may represent linked *trans*-acting factors that regulate ELPs. To examine these effects, we intend to generate a doubled haploid (DH) population from a cross between isolines that can be used to map ELPs and SFPs in a genetical genomics approach to determine eQTL associated with the *Yr5* phenotype. A DH population will be used as ELPs and SFPs are dominant markers that cannot be used accurately in a population with high levels of heterozygosity. Also, we intend to construct a custom microarray containing only the ELP and SFP-containing probe sets identified in this study, which will be used to map these ELPs and SFPs in a more cost-effective manner. Although West et al. (2007) found that transcript level variation in the parents of a progeny underestimates the variation present in that progeny, most large effect *cis*-eQTL can be still determined using only the parental transcript variation. Subsequently, the use of a focused microarray may be useful for associating *cis*-eQTL with the*Yr5* phenotype, and may enable the identification of regulatory influences of putative linked *trans*-acting SFPs on ELPs. Results of such a study may identify particular transcripts and regions of the genome important for regulating the expression of the agriculturally important *Yr5* phenotype.

Acknowledgments This research was supported in part by the US Department of Agriculture (USDA), Agricultural Research Service (ARS) (Project No. 5348-22000-014-00D), USDA-ARS Postdoctoral Program, and Washington Wheat Commission (Project No. 13C-3061-3923). PPNS No. 0470, Department of Plant Pathology, College of Agricultural, Human, and Natural Resources Research Center, Project No. WNP00823. The authors acknowledge Derek Pouchnik (Washington State University, School of Molecular Biosciences) for carrying out GeneChip hybridizations and scanning. We thank Dr. Scot Hulbert and Dr. Patricia Okubara for their critical review of the manuscript.

References

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on bias and variance. Bioinformatics 19:185–193

Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J (2003) Large-scale identification of single-feature polymorphisms in complex genomes genome res 13:513–523

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FP, Kim IF, Markowitz V, Matesse JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M (2001) Minimum information about a microarray experiment (MIAME)-towards standards for microarray data. Nat Genet 29:365–371

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between *Arabidopsis* FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. Proc Natl Acad Sci USA 101:15670–15675

Chen XM, Line RF, Shi ZX, Leung H (1998) Genetics of wheat resistance to stripe rust. In: Slinkard A (ed) 9th international wheat genetics symposium. University Extension Press, Saskatoon, pp 237–239

Chen XM, Soria MA, Yan GP, Sun J, Dubcovsky J (2003) Development of sequence tagged site and cleaved amplified polymorphic sequence markers for wheat stripe rust resistance gene *Yr5*. Crop Sci 43:2058–2064

Chen WJ, Chang SH, Hudson ME, Kwan W-K, Li J, Estes B, Knoll D, Shi L, Zhu T (2005) Contribution of transcriptional regulation to natural variations in *Arabidopsis*. Genome Biol 6:R32

Cong B, Liu JP, Tanksley S (2002) Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. Proc Natl Acad Sci USA 99:13606–13611

Coram TE, Wang MN, Chen XM (2008) Transcriptome analysis of the wheat-*Puccinia striiformis* f sp. tritici interaction. Mol Plant Pathol 9:157–169

Cui X, Xu J, Asghar R, Condamine P, Svensson JT, Wanamaker S, Stein N, Roose M, Close TJ (2005) Detecting single-feature polymorphisms using oligonucleotide arrays and robustified projection pursuit. Bioinformatics 21:3852–3858

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3:43–52

Gassmann W, Hinsch ME, Staskawicz BJ (1999) The *Arabidopsis RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. Plant J 20:265–277

Gentleman R, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW, Dangl JL (1995) Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. Science 269:843–846

Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP (2003) Summaries of affymetrix GeneChip probe level data. Nucleic Acids Res 31:e15

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25

Jayasinghe R, Kong S, Coram TE, Kaganovitch J, Xue CCL, Li CG, Pang ECK (2007) Construction and validation of a prototype microarray for efficient and high-throughput genotyping of angiosperms. Plant Biotechnol J 5:282–289

Jordan MC, Somers DJ, Banks TW (2007) Identifying regions of the wheat genome controlling seed development by mapping expression quantitative trait loci. Plant Biotechnol J 5:442–453

Kliebenstein DJ, Lambrix V, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in *Arabidopsis thaliana*. Plant Cell 13:681–693

Kliebenstein DJ, West MA, Van Leeuwen H, Kim K, Doerge RW, Michelmore RW, St. Clair DA (2006a) Genomic survey of gene expression diversity in *Arabidopsis thaliana*. Genetics 172:1179–1189

Kliebenstein DJ, West MA, Van Leeuwen H, Loudet O, Doerge RW, St. Clair DA (2006b) Identification of QTLs controlling gene expression networks defined *a priori*. BMC Bioinformatics 7:308

Lambrix V, Reichelt M, Mitchell-Olds T, Kliebenstein DJ, Gershenzon J (2001) The *Arabidopsis* epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. Plant Cell 13:2793–2807

Law CN (1976) Genetic control of yellow rust resistance in *T. spelta album*. Plant Breeding Institute, Cambridge, Annual Report 1975, pp 108–109

Ling P, Chen XM (2005) Construction of a hexaploid wheat (*Triticum aestivum* L.) bacterial artificial chromosome library for cloning genes for stripe rust resistance. Genome 48:1028–1036

Lupton FGH, Macer RCF (1962) Inheritance of resistance to yellow rust (*Puccinia glumarum* Erikss & Henn) in seven varieties of wheat. Trans Br Mycol Soc 45:21–45

Macer RCF (1966) The formal and monosomic genetic analysis of stripe rust (*Puccinia striiformis*) resistance in wheat. In: Mackey J (ed) 2nd international wheat genetics symposium. Hereditas Suppl, Lund, pp 127–142

McIntosh RA, Hart GE, Devos KM, Gale MD, Rogers WJ (1998) Catalogue of gene symbols for wheat. In: Slinkard A (ed) 9th international wheat genetics symposium. University Extension Press, Saskatoon, pp 1–235

Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvorak J, Linkiewicz AM, Ratnasiri A, Dubcovsky J, Bermudez-Kandianis CE, Greene RA, Kantety R, La Rota CM, Munkvold JD, Sorrells SF, Sorrells ME, Dilbirligi M, Sidhu D, Erayman M, Randhawa HS, Sandhu D, Bondareva SN, Gill KS, Mahmoud AA, Ma X-F, Gustafson JP, Miftahudin, Conley EJ, Nduati V, Gonzalez-Hernandez JL, Anderson JA, Peng JH, Lapitan NLV, Hossain KG, Kalavacharla V, Kianian SF, Pathan MS, Zhang DS, Nguyen HT, Choi D-W, Fenton RD, Close TJ, McGuire PE, Qualset CO, Gill BS (2004) A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics 168:701–712

Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J, Cardle L, Marshall DF, Waugh R (2005) Single-feature polymorphism discovery in the barley transcriptome. Genome Biol 6:R54

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–302

Shen L, Gong J, Caldo RA, Nettleton D, Cook D, Wise RP, Dickerson JA (2005) BarleyBase—an expression profiling database for plant genomics. Nucleic Acids Res 33:D614–D618

Smith PH, Hadfield J, Hart NJ, Koebner RMD, Boyd LA (2007) STS markers for the wheat yellow rust resistance gene *Yr5* suggest a NBS-LLR-type resistance gene cluster. Genome 50:259–265

Smyth GK (2005) Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (eds) Bioinformatics and computational biology solutions using r and bioconductor. Springer, New York, pp 397–420

Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson PJ, Qi LL, Echalier B, Gill BS, Matthews DE, Lazo GR, Chao S, Anderson OD, Edwards H, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Zhang D, Nguyen HT, Peng J, Lapitan NLV, Gonzalez-Hernandez JL, Anderson JA, Hossain K, Kalavacharla V, Kianian SF, Choi DW, Close TJ, Dilbirligi M, Gill KS, Steber C, Walker-Simmons MK, McGuire PE, Qualset CO (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13:1818–1827

R Development Core Team (2006) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98:5116–5121

Van Leeuwen H, Kliebenstein DJ, West MA, Kim K, Van Poecke R, Katagiri F, Michelmore RW, Doerge RW, St. Clair DA (2007) Natural variation among *Arabidopsis thaliana* accessions for transcriptome response to exogenous salicylic acid. Plant Cell 19:2099–2110

Walia H, Wilson C, Condamine P, Ismail AM, Xu J, Cui X, Close TJ (2007) Array-based genotyping and expression analysis of barley cv Maythorpe and Golden Promise. BMC Genomics 8:87

Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. Nature 398:236–239

Wellings CR, Singh RP, McIntosh RA, Pretorius ZA (2004) The development and application of near isogenic lines for the wheat stripe (yellow) rust pathosystem. 11th international cereal rusts and powdery mildew conference. John Innes Centre, Norwich, p 39

West MA, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, St. Clair DA, Michelmore RW (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. Genome Res 16:787–795

West MA, Kim K, Kliebenstein DJ, Van Leeuwen H, Michelmore RW, Doerge RW, St. Clair DA (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. Genetics 175:1441–1450

Yan GP, Chen XM, Line RF, Wellings CR (2003) Resistance gene-analog polymorphism markers co-segregating with the *Yr5* gene for resistance to wheat stripe rust. Theor Appl Genet 106:636–643

Yuan Q, Ouyang S, Liu J, Suh B, Cheung F, Sultana R, Lee D, Quackenbush J, Buell CR (2003) The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. Nucleic Acids Res 31:229–233

Zhu T, Salmeron JM (2007) High-definition genome profiling for genetic marker discovery. Trends Plant Sci 12:196–202