Systems Genetics of Environmental Response in the Mature Wheat Embryo

Jesse D. Munkvold,*,[1] Debbie Laudencia-Chingcuanco,[†] and Mark E. Sorrells*,[2]
*Department of Plant Breeding and Genetics, College of Agriculture and Life Sciences, Cornell University, Ithaca, New York 14853, and [†]Genomics and Gene Discovery Unit, U.S. Department of Agriculture–Agricultural Research Service, Western Regional Research Center, Albany, California 94710

ABSTRACT Quantitative phenotypic traits are influenced by genetic and environmental variables as well as the interaction between the two. Underlying genetic × environment interaction is the influence that the surrounding environment exerts on gene expression. Perturbation of gene expression by environmental factors manifests itself in alterations to gene co-expression networks and ultimately in phenotypic plasticity. Comparative gene co-expression networks have been used to uncover biological mechanisms that differentiate tissues or other biological factors. In this study, we have extended consensus and differential Weighted Gene Co-Expression Network Analysis to compare the influence of different growing environments on gene co-expression in the mature wheat (*Triticum aestivum*) embryo. This network approach was combined with mapping of individual gene expression QTL to examine the genetic control of environmentally static and variable gene expression. The approach is useful for gene expression experiments containing multiple environments and allowed for the identification of specific gene co-expression modules responsive to environmental factors. This procedure identified conserved coregulation of gene expression between environments related to basic developmental and cellular functions, including protein localization and catabolism, vesicle composition/trafficking, Golgi transport, and polysaccharide metabolism among others. Environmentally unique modules were found to contain genes with predicted functions in responding to abiotic and biotic environmental variables. These findings represent the first report using consensus and differential Weighted Gene Co-expression Network Analysis to characterize the influence of environment on coordinated transcriptional regulation.

QUANTITATIVE phenotypic traits are governed by genetics, environment, and interactions between the two (G×E). The extent of G×E for a particular trait in part determines heritability and our understanding of the underlying biology. Transcriptional regulation of genes plays a large role in the response of an organism to the environment, and this modulation of gene expression is itself genetically controlled and subject to G×E interactions (Landry *et al.* 2006; Li *et al.* 2006; Smith and Kruglyak 2008). A number of studies have used broad transcriptional profiling in segregating or natural populations to identify genetic loci that control gene expression, commonly referred to as genetical genomics or gene expression QTL (eQTL) mapping (Jansen and Nap 2001; Brem *et al.* 2002; Schadt *et al.* 2003; Brem and Kruglyak 2005; Zhang *et al.* 2011). In plants a number of studies have been conducted on a range of species including *Arabidopsis thaliana*, barley (*Hordeum vulgare*), maize (*Zea mays*), poplar (*Populus* spp.), and wheat (*Triticum aestivum*) (Decook *et al.* 2006; Jordan *et al.* 2007; Keurentjes *et al.* 2007; West *et al.* 2007; Potokina *et al.* 2008; Drost *et al.* 2010; Druka *et al.* 2010; Holloway *et al.* 2011; Zhang *et al.* 2011). Using a genetical genomics approach, gene expression by environment interaction has been studied in *Caenorhabditis elegans* and yeast, but to date no studies in plants have taken advantage of gene expression measurements in more than one environment to help in understanding G×E (Li *et al.* 2006; Smith and Kruglyak 2008).

Only recently have researchers taken advantage of systems biology and network approaches developed in yeast and animal systems to help to understand the connections between

doi: 10.1534/genetics.113.150052
Manuscript received February 2, 2013; accepted for publication February 26, 2013
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150052/-/DC1.
The complete gene expression data are available from NCBI's Gene Expression Omnibus under accession nos. GPL15156, GSE35300, and GSE35407 (http://www.ncbi.nlm.nih.gov/geo/).
[1]Present address: Applied Systems Biology, Keygene Inc., Rockville Innovation Center, 155 Gibbs St., Rockville, MD 20850.
[2]Corresponding author: Department of Plant Breeding and Genetics, College of Agriculture and Life Sciences, 240 Emerson Hall, Cornell University, Ithaca, NY 14853. E-mail: mes12@cornell.edu

global gene expression and biological mechanisms in plants (Keurentjes *et al.* 2007; Mounet *et al.* 2009; Lee *et al.* 2010; Bassel *et al.* 2011; Zhang *et al.* 2011). These naive systems approaches that do not require prior information regarding gene function or metabolic pathways are of great potential for characterizing biological systems in plant species such as wheat (*T. aestivum*), which has limited genomic resources. Additional strategies have emerged for comparing gene expression networks between tissues, sex, or species of interest (Oldham *et al.* 2006; Fuller *et al.* 2007; Langfelder and Horvath 2007; Langfelder *et al.* 2011). These consensus and differential network analysis approaches allow researchers to use multiple variables to differentiate gene co-expression patterns of interest that are conserved and unique with regards to the variables of interest.

Here we present work detailing the transcriptional architecture and gene networks in the mature wheat embryo across ~100 lines from a doubled-haploid population sampled in two representative field environments. The gene expression data were used to map eQTL for individual environments, which were used to identify both conserved and divergent patterns of genetic regulation of transcript abundance. The gene expression data were also used to construct single-environment Weighted Gene Co-Expression Networks (WGCNA) that could be related back to patterns of transcriptional regulation. To study environment-specific co-expression, a differential network analysis approach was utilized. The environmentally shared and unique co-expression modules were analyzed for biological relevance. This study forms the basis upon which detailed relationships can be discovered between important phenotypic seed traits in wheat and how the underlying gene co-expression networks are influenced by genetics and the environment. The complete gene expression data are available from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under the nos. GPL15156, GSE35300, and GSE35407.

Materials and Methods

Plant material and embryo tissue harvest

The double-haploid lines used in this study, previously described in Munkvold *et al.* (2009), were created from a cross between two elite soft white winter wheat varieties that are both well adapted to the northeastern United States but differ in preharvest sprouting (PHS) resistance: Cayuga (PHS-resistant) and Caledonia (PHS-susceptible). Plant materials from 2 of the 16 environments described in Munkvold *et al.* (2009), Helfer 2004 (designated as H04) and Ketola 2005 (designated as K05), were used for gene expression analysis (Supporting Information, File S1). Both environments are located near Ithaca, New York, and do not differ in photoperiod; however, the environments have different soil types and experienced temperature and rainfall differences due to year effects (Figure S1). Seed of the Cayuga × Caledonia doubled-haploid population is available upon request from the corresponding author.

For each line, five spikes were harvested at physiological maturity as evidenced by loss of chlorophyll from the glumes. The spikes were dried for 5 days at ambient temperature and humidity. Approximately 20 seeds from the middle two-thirds of each spike at the primary and secondary spikelet positions were harvested by hand. Embryos were collected by hand dissection, snap-frozen in liquid nitrogen, and stored at −80° until used for RNA isolation.

Embryo RNA isolation

Total RNA from wheat embryos was isolated using a modified TRIZOL (Invitrogen, Carlsbad, CA) extraction protocol. To remove DNA contamination, the extracted RNA was treated with RQ1 DNAse (Promega, Madison, WI) according to the manufacturer's protocol. The RNA was further purified with a Qiagen RNeasy column using the RNA cleaning protocol (Qiagen, Valencia, CA). Total RNA concentration and purity was measured by spectrophotometry at OD_{260} and OD_{280}. Total RNA integrity was checked by separation of 500 ng on a 1% agarose gel and evaluated for distinct 28S and 18S rRNA bands with no smearing. The wheat embryo total RNA samples were stored at −80°.

Target labeling

For the H04 population, RNA reverse transcription and target labeling was done using a modified protocol based on protocols from Microarrays Inc. (http://www.microarrays.com), TIGR (http://compbio.dfci.harvard.edu/compbio/research/labprotocols), and U. S. Department of Agriculture–Agricultural Research Service (USDA–ARS) Western Regional Research Center, Albany, CA; (http://www.ars.usda.gov/main/site_main.htm?modecode=53-25-00-00). The resulting purified amine-modified complementary DNA (cDNA) was dried using vacuum centrifugation at 25° and stored at −80° prior to conjugation with Alexa 555 or Alexa 647 dyes (Invitrogen). Following dye coupling, the dye incorporation was calculated as described by Bowtell and Sambrook (2003). Any samples with less than one molecule of dye incorporation per 50 bp of cDNA were discarded. The experimental and reference samples to be hybridized on the same microarray were mixed together with balanced amounts of each dye. The labeled samples were dried using a SpeedVac and placed at −80°. To address issues with dye incorporation and signal intensity encountered for the H04 location, embryo RNA from the K05 population was amplified and labeled using the Amino Allyl MessageAmp II kit with accompanying protocols (Applied Biosystems/Ambion, Austin, TX).

Microarray gene expression analysis

For the microarray analysis of gene expression, a reference experimental design was used with complete dye-swap technical replication. Each experimental sample was hybridized against a pre-anthesis spike RNA sample from the same pooled RNA source.

The oligonucleotide microarrays consisted of 70-bp probes from 17,729 unique wheat suequences and 50

controls spotted in duplicate as previously described (Pumphrey *et al.* 2009). The design was done in a collaborative effort with the Genome Canada/Genome Prairie/Genome Quebec program for Functional Genomics of Abiotic Stress, the USDA–ARS, Genomics and Gene Discovery Research Unit (Albany, CA), and the Australian Centre Plant Functional Genomics (Adelaide, NSW, Australia). All prehybridization, hybridization, and washing steps were carried out by hand using reagents from the Pronto! Universal Hybridization Kit with accompanying protocols (Corning Life Sciences, Acton, MA).

The hybridized microarray slides were scanned using an Axon 4000B microarray scanner, and microarray image analysis was done using the Genepix 6.1 software program (Molecular Devices, Union City, CA). Background-corrected measures of gene expression were normalized block-wise within each slide using LOWESS smoothing and variance scale normalization as implemented in the TM4:Midas software program (http://www.tm4.org/midas.html). Between-slide variance scale normalization was also conducted to facilitate the comparison of gene expression (Quackenbush 2002; Yang *et al.* 2002; Saeed *et al.* 2003). All spots that were flagged absent or corrupted were removed, and the average of the in-slide and dye-swap replicates was used for the final gene expression measure. All cases of oligo by dye bias were also removed before calculating the final gene expression measures (File S2).

eQTL mapping

The eQTL analysis was done using the QTLCartographer command line software program implemented on an Apple X-serve cluster (Basten *et al.* 1994, 2002). Initial analyses were done using the SRMapQTL function model 2 with forward stepwise regression and backward elimination thresholds at $P < 0.01$. These data were used in the ZmapQTL function for composite interval mapping (CIM) with a window of 10 cM, 2-cM steps, and background marker selection model 6. The ZmapQTL results were then summarized using the QTLcartographer Eqtl function with a likelihood ratio (LR) cutoff of 11.5. All features with at least one eQTL with an LR value >11.5 were used for permutation testing to set an appropriate experiment-wise significance threshold. Automation of permutation testing was done using a custom Perl script to run the QTL Cartographer permutation tool permute.pl with 1000 permutations and the same settings as those used for the initial CIM analysis (File S3 and File S4). The gene-specific LR value at the $P < 0.05$ threshold was used as the significance threshold for each gene expression trait.

Cis/trans regulation

To determine if the observed eQTL were due to *cis*- or *trans*-regulation, the full sequence from which the long oligos were designed was compared to the wheat deletion bin-mapped EST sequences using BLAST (http://wheat.pw.usda.gov/GG2/blast.shtml). A bit score of >300 was used as a significance threshold to identify the best deletion bin-mapped EST match. The known EST genome locations were then compared to the eQTL locations and assigned *cis*-regulation if the EST and eQTL were located on the same chromosome.

WGCNA

The full gene expression dataset used for eQTL analysis from each environment was used as input for single-environment WGCNA using the WGCNA package written for the statistical analysis program R (http://cran.r-project.org; Langfelder and Horvath 2008). A soft-thresholding power of $\beta = 8$ was chosen for calculating adjacency by evaluating the scale-free topology model fit as discussed in Zhang and Horvath (2005). Due to computational limitations, initial network construction was done block-wise in sets of 5000 gene expression traits using the auto-network construction function. Genes that were unconnected in the resulting network were removed from the input data, and network construction was rerun using single-block manual construction. The module eigengene clustering dendrogram was evaluated, and similar modules were manually merged. Node and edge files were generated for use with the Cytoscape network visualization program (Shannon *et al.* 2003; http://www.cytoscape.org).

Consensus network construction was conducted using the trimmed datasets established in the single-environment WGCNA. Network construction and module detection used the one-step automatic method with a soft threshold of $\beta = 8$ and a minimum module size of 20.

Analyses and comparisons of the eigengene network and consensus modules were done as described in the WGCNA documentation (http://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/) (File S6).

Detection of module eQTL hotspots

Transcriptional regulation hotspots for environment and consensus modules were detected by comparing the number of eQTL from a module to the number of eQTL from non-module and non-network genes using a Fisher's exact test. Tests were conducted in 2-cM bins for all bins containing one or more eQTL from the module of interest. Significance thresholds were set by calculating $\alpha < 0.05/n$, where n was the number of tests conducted for the module (File S7).

Gene-set enrichment analysis

Due to the limited availability of genome sequence and gene annotation in wheat, the sequence used for designing oligonucleotide probes was compared to Arabidopsis proteins using BLASTx. The best hit from Arabidopsis with a bit score >300 and an expectation value (E-value) $<10^{-5}$ was used as the representative gene for gene-set enrichment analysis. Gene sets corresponding to individual environment and consensus modules were compared against the entire oligonucleotide array as background using the DAVID functional annotation tool (Huang Da *et al.* 2009a,b). Interesting

Table 1 Summary of the genome-wide eQTL results over two environments

Environment	Lines with expression data	Genes with expression data	eQTL	Genes with eQTL	Genes with shared or unshared eQTL	Shared eQTL	*Cis*-regulation	*Trans*-regulation
HO4	96	14,288	2795	2104	401	109	14.3% (65/454)	85.7% (389/454)
K05	115	11,745	3336	2455	401	109	16% (108/676)	84% (568/676)

Parentheses indicate the number of eQTL out of the total that could be assigned either *cis* or *trans* regulation.

functional annotation groups were identified using the functional annotation clustering tool with high stringency and a focus on clusters with an EASE score >1.3.

Results

eQTL mapping and distribution

Gene expression data from 96 H04 and 115 K05 double-haploid lines were used for eQTL mapping. Of the 17,729 genes on the array, 14,288 and 11,745 genes, for H04 and K05, respectively, had sufficient expression to be used for further analyses. The gene expression measurements for the H04 and K05 environments were used as gene expression traits in CIM QTL mapping with individual-gene experiment-wise significance thresholds based on 1,000 permutations ($P < 0.05$). For the H04 location, 2795 eQTL from 2104 gene expression traits were identified (Table 1). Single eQTL explained 7–70% (mean 15%) of individual transcript variation in the H04 environment and 2–80% (mean 14%) of individual transcript variation in the K05 environment. Of the 2104 expression traits with at least one eQTL in the H04 environment, 41 had multiple eQTL that mapped to homeologous chromosomes. For the K05 environment, 3336 eQTL were mapped for 2455 expression traits, with 84 exhibiting multiple eQTL on homeologous chromosomes (Table 1). No evidence for an allelic bias in the control of gene expression was observed for either environment. In the H04 environment the allelic contribution to increased gene expression was 51% from maternal alleles to 49% from paternal alleles, and in the K05 environment the same comparison was 47% maternal to 53% paternal.

The eQTL were distributed on all linkage groups covering the 21 wheat chromosomes (Figure 1). Accumulations of eQTL at individual loci could be observed across the genome. Frequently, eQTL accumulated near the ends of linkage groups, most notably on chromosomes 6D (bin 632) and 7D (bins 739 and 743) in the H04 environment; 2B (bin 178), 5D (bin 551), and 7D (bin 648) in the K05 environment; and 1D (bin 47) in both environments (Figure 1).

Cis/trans analysis

To estimate *cis/trans* regulation for the corresponding eQTL, the putative chromosomal locations of the genes represented on the array were identified by comparing the target sequences used to design the long oligos with the deletion bin-mapped ESTs from wheat. A total of 454 and 676 eQTL, from H04 and K05, respectively, were able to be assigned *cis* or *trans* regulation. The two environments were very similar with *cis*-regulation rates of 14.3% for H04 and 16% for K05 (Table 1). The average percentage of the transcript variation explained (r^2) by *cis* QTL was 18.2% in the H04 environment and 17.6% in the K05 environment, which was found to be significantly greater ($P < 0.01$) than the average r^2 values of 14.8 and 13.6% for *trans* eQTL in each environment, respectively.

eQTL comparison of environments

Between the two environments 401 expression traits had an eQTL in both environments, of which 109 had an eQTL within 20 cM (Table 1 and Figure 2). Of the 109 eQTL with a shared location between environments, 20 were among the 50 most significant eQTL in both environments. Of the 20 most significant shared eQTL, 5 could be assigned a *cis/trans* relationship with 4 of the 5 found to be in *cis*. Several small clusters of genes were found to have colocated eQTL in the H04 environment and maintained colocated eQTL on a different chromosome in the K05 environment (Figure 2). Genes with eQTL in different locations between the environments suggest G×E interaction in the control of gene expression.

Individual environment WGCNA

Mapping eQTL for each gene expression trait allows for general inferences regarding patterns of transcriptional control across the genome. However, this type of analysis is of limited utility for linking transcriptional regulators to important biological processes. A WGCNA approach was used to identify groups of genes (modules) with coordinated expression. Hierarchical clustering of the modules based on the module eigengene suggested higher relatedness between pairs or groups of modules, *e.g.*, high adjacency between H04-brown and H04-midnight blue and between K05-pink and K05-yellow (Figure 3).

To identify hotspots for transcriptional control of modules from the individual environments, we compared the number of eQTL in individual modules to the number of eQTL in other modules or unconnected in the network at 2-cM intervals along the linkage map using the Fisher's exact test (corrected $P < 0.05$). This analysis identified 11 eQTL hotspots from nine modules in the H04 location and 12 eQTL hotspots from nine modules in the K05 environment (Figure 1).

Comparison of networks across environments

To compare gene co-expression modules between environments, consensus WGCNA was conducted using 4739 gene expression traits given a single-environment module assignment in one or both of the environments. The consensus

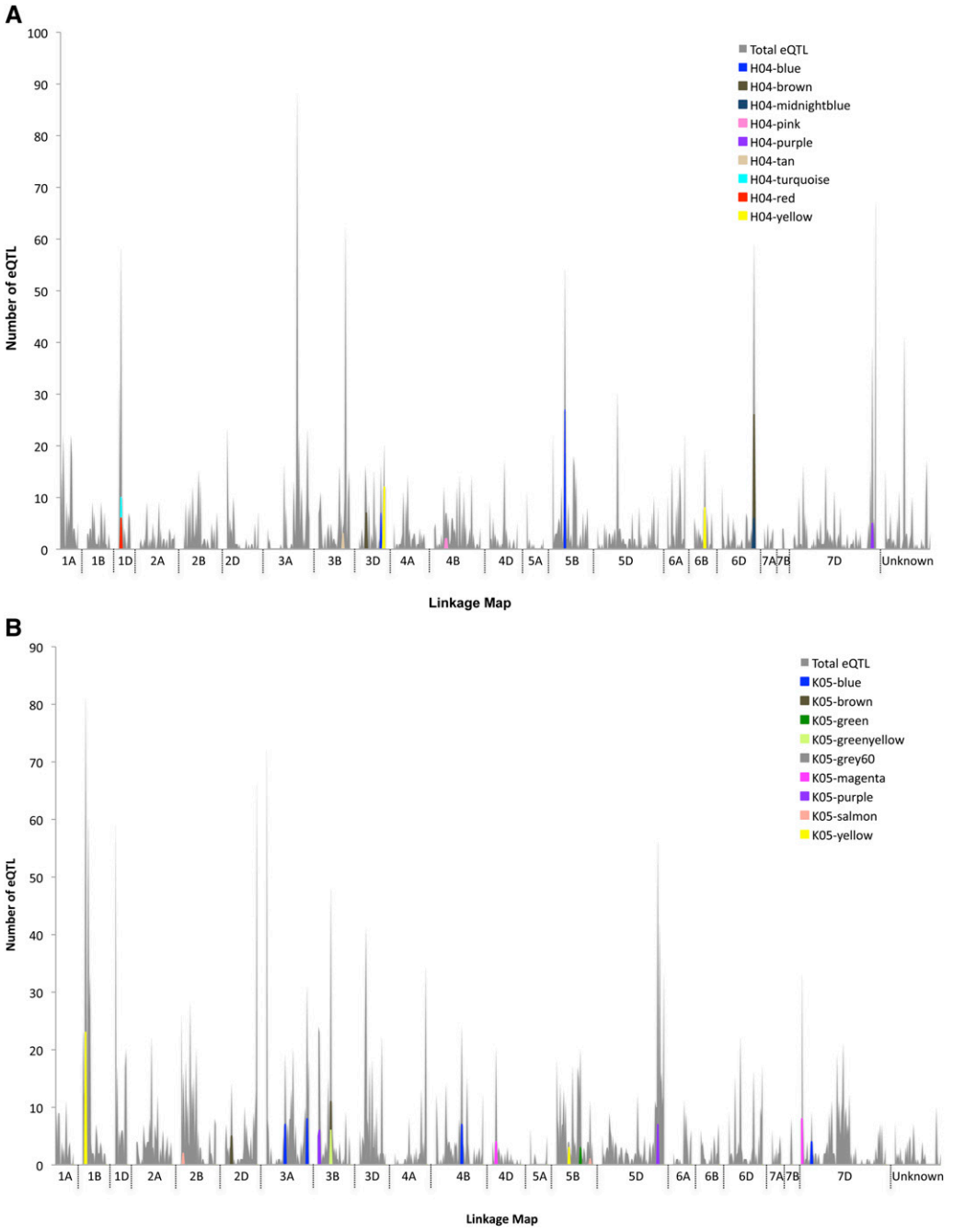

Figure 1 Distribution of eQTL along the C×C linkage map in 2-cM bins. (A) Colored peaks represent significantly enriched loci corresponding to H04 environment-specific WGCNA modules. Enrichment was determined by comparing the number of module genes with eQTL at a locus with the number of eQTL from non-module genes using Fisher's exact test and a Bonferroni correction of $P < 0.05$. (B) Distribution and module enrichment plot for the K05 environment specific co-expression modules.

network identified 10 modules containing 1312 genes. These modules represent groups of gene expression traits that are co-expressed in both the H04 and K05 environments. Network preservation figures modeled on those of Langfelder and Horvath (2007) can be found in Figure 4. Clustering based on the environment-specific eigengene identified a single meta-module composed of the consensus modules Con-brown and Con-yellow. These consensus modules cluster together when using module eigengenes calculated from the H04 data (Figure 4, A and C) or from the K05 data (Figure 4, B and F). The overall network preservation (D) was found to be 0.76 (Figure 4D) with generally high correspondence between the module eigengenes when calculated from the H04 or K05 data (Figure 4E).

To explore the relationship between individual environment and consensus modules, we identified the number of genes in each single-environment module related to their module assignment in the consensus network. A Fisher's exact test was then used to determine the significance of the correspondence between the two modules (Figure 5). This comparison identified seven H04-specific modules: H04-yellow, H04-brown, H04-midnight blue, H04-tan, H04-lightcyan, H04-greenyellow, and H04-cyan. Comparison of the H04 WGCNA also uncovered a number of significantly

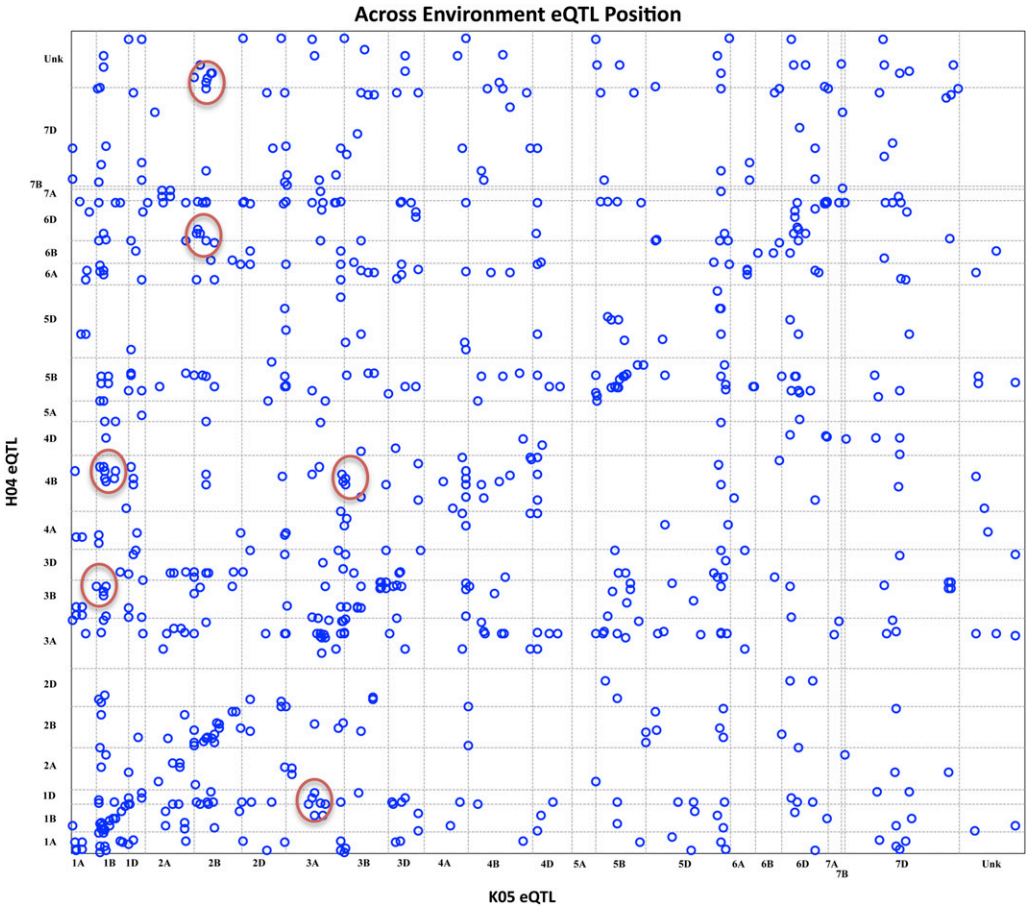

Figure 2 Dot-plot comparison of eQTL positions from genes with eQTL in both the H04 and K05 environments. Axes follow the C×C linkage map with the H04 environment on the *y*-axis and the K05 environment on the *x*-axis. Points along the middle diagonal represent eQTL with a shared position in each environment. Red circles indicate loci where genes with colocated eQTL in one environment colocate at a different genome location in the second environment.

preserved modules (*e.g.*, H04-blue with Con-turquoise and H04-pink with Con-black). The H04-turquoise module, which contained 769 genes, was significantly split among six different consensus modules with more than one-half of the genes not contained in the network (Con-gray) (Figure 5A).

A comparison between the K05 environment WGCNA and the consensus identified 13 single-environment modules unpreserved in the consensus network. Most notably, the K05-yellow, K05-pink, and K05-brown modules contained >100 genes each. The K05-blue module, which contained 503 genes, was significantly preserved among five different consensus modules with 204 genes falling out of the consensus network (Con-gray). Almost 50% of the 748 genes found in the K05-turquoise module were preserved in the Con-turquoise module (Figure 5B).

Biological significance of environment-specific modules

The environment-specific modules represent groups of genes that are coordinately regulated in only a single environment (Figure 6). To better understand the biological significance of these modules, we conducted gene ontology (GO) enrichment analysis using the DAVID functional annotation tool (http://david.abcc.ncifcrf.gov). We used functional annotation clustering and focused on annotation clusters with an EASE score >1.3.

The H04-brown module contained 223 genes that were unconnected in the consensus network. Functional analysis of those genes identified annotation clusters related to protein kinase activity, monooxygenase activity, calcium binding, ATP binding, and toxin catabolism. The most significantly enriched GO biological process (GO-BP) category was related to a response to chitin that included two WRKY transcription factors, WRKY33 (AT2G38470) and WRKY40 (AT1G80840), known to be involved in resistance to fungal pathogens (Xu *et al.* 2006; Zheng *et al.* 2006). Enrichment for phosphorylation activity was also identified, including three calcium-dependent kinases, five uncharacterized receptor-like kinases, and a MAPKKK. Taken together, the functional annotations suggest that the H04-brown module represents an environmentally specific response to pathogens.

The H04-yellow module represents a second large environment-specific module. Of the 177 module genes, 175 were not connected in the consensus network. Functional annotation analysis revealed clusters related to lipase/esterase activity, transcription factors, and hormone response. The most significantly enriched GO-BP category was a response to wounding, followed by categories related to regulation of transcription and hormonal response. These enriched GO categories were composed of genes encoding proteins related to jasmonic acid (JA) response, including a known negative regulator of defense response ATAF1 (AT1G01720) (Wang *et al.*

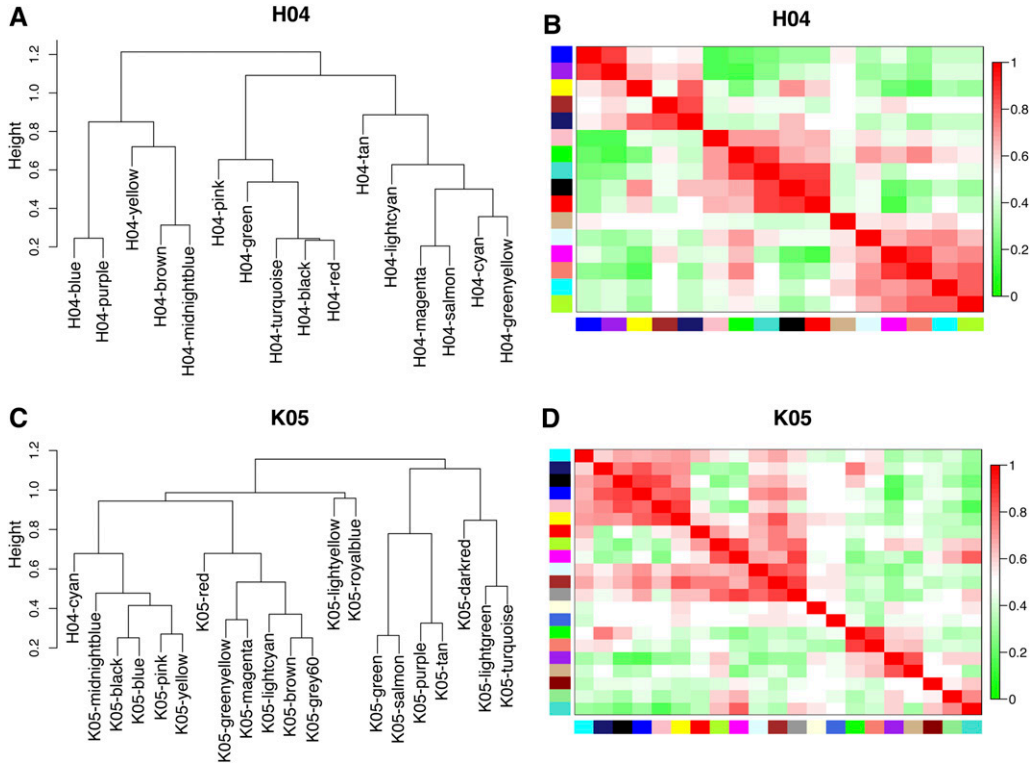

Figure 3 Relationship between single-environment WGCNA module eigengenes. (A) Hierarchical clustering tree based on H04 single-environment WGCNA module eigengenes. (B) Heat map of adjacency between H04 single-environment modules. Colors are from low adjacency (green) to high adjacency (red). (C) Eigengene clustering tree based on K05 single-environment modules. (D) Corresponding adjacency heat map from K05 single-environment modules. Colors are from low adjacency (green) to high adjacency (red).

2009), the JAZ repressor protein JAZ2 (AT1G74950) (Thines *et al.* 2007), a negative regulator of JA-induced transcription AtERF4 (AT3G15210) (McGrath *et al.* 2005), and the JA-induced transcription factor MYC2 (AT1G32640) (Lorenzo *et al.* 2004). The functional annotation and GO enrichment suggests that the H04-yellow module may also represent an environmentally specific response to biotic stress.

A total of 216 of the 220 genes in the K05-yellow module and 116 of the 119 genes in the K05-pink module were not connected in the consensus network. Based on the clustering of the K05 module eigengenes (Figure 3B), the K05-yellow and K05-pink modules were combined and used for functional annotation clustering. The DAVID functional annotation tool identified significant clusters (EASE $>$ 1.3) related to heat-shock and peroxidase activity.

Similar to the K05-pink and K05-yellow modules, the K05-brown, K05-grey60, K05-lightcyan, and K05-green/yellow modules were not conserved in the consensus network and are closely grouped based on eigengene clustering. Functional annotation clustering of the genes from these four modules identified significant clusters related to protein phosphatase 2C proteins and heat-shock proteins. Similar analyses of the K05-purple and K05-tan module genes resulted in significant clusters related to the splicesome complex and bZIP transcription factors.

Genetic control of environment-specific modules

To determine the genetic control of H04-specific modules, the number of module genes with eQTL in each linkage bin was compared to the number of non-module gene eQTL using the Fisher's exact test. A significant eQTL enrichment of H04-brown module genes was observed on chromosome 6D (bin 633), which also colocated with several eQTL from H04-midnight blue and H04-yellow module genes (Figure 1A). These 54 genes were highly interconnected and contained eight genes that encode for proteins with known roles in disease resistance (Figure 7A). Additionally, the H04-brown module was significantly enriched for eQTL on chromosome 3D (bin 273). Of these 21 genes, 5 encode proteins with a possible role in disease resistance, including the powdery mildew resistance proteins MLO2 (AT1G11310) and MLO12 (AT2G39200) (Consonni *et al.* 2010); NHL10 (AT2G35980), which is highly induced during the hypersensitive response (Zheng *et al.* 2004); phenylalanine ammonia-lyase PAL1 (AT2G37040), which may play a role in basal defense (Mishina and Zeier 2007); and the enzyme cinnimate 4- hydroxylase C4H (AT2G30490), which is immediately downstream from PAL1 at the start of the general phenylpropanoid pathway. Additionally, five uncharacterized protein kinase genes were also regulated by the locus on chromosome 3D, including the calcium-dependent protein kinases CPK7 (AT5G12480) and CPK28 (AT5G66210) (Figure 7B).

A K05-yellow module eQTL hotspot was detected on chromosome 1B (bin 24), which also colocated with eQTL for seven genes in the pink module. The most highly connected among these coregulated genes was the well-characterized gene encoding the RNA-binding Slicer homolog AGO1 (AT1G4841), which is involved in RNA silencing (Baumberger and Baulcombe 2005). Additionally, six other

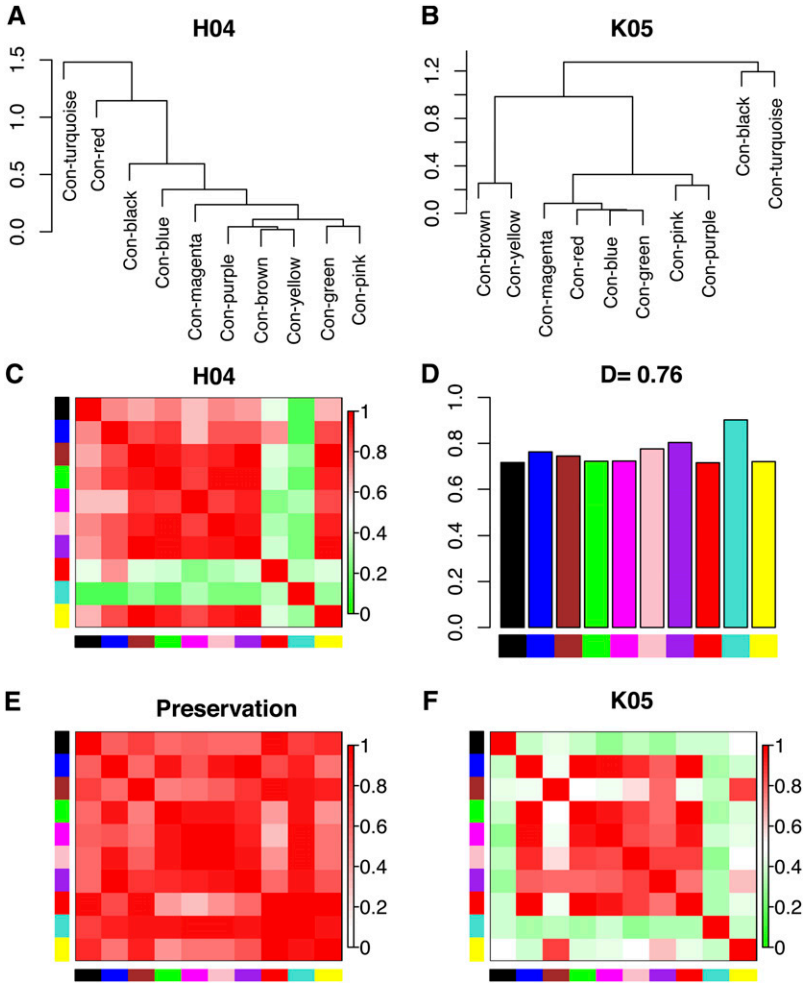

Figure 4 Comparison of consensus eigengene networks for the H04 and K05 environments based on figures in Langfelder and Horvath (2007). (A) Hierarchical clustering of consensus modules based on eigengenes from the H04 environment. (B) Hierarchical clustering of consensus modules based on eigengenes from the K05 environment. (C) A heat-map relationship between consensus modules based on adjacency for the H04 environment. Colors are from low adjacency (green) to high adjacency (red). (D) Preservation of consensus module eigengenes between the H04 and K05 environments. The preservation of the entire network (D) was determined to be 0.76. (E) Heat map of the preservation network ranging from low adjacency (white) to high adjacency (red). (F) Hierarchical clustering heat map for the K05 environment as shown in C for the H04 environment.

genes that code for proteins with RNA-binding domains were regulated by the 1B locus (Figure 7C). Functional annotation also revealed genes related to abiotic stress, including homologs to an E3 ubiquitin ligase SAP5 (AT3G12630) (Kang *et al.* 2011), the urea transporter DUR3 (AT5G45380) (Liu *et al.* 2003; Kojima *et al.* 2007), and the glutathione *S*-transferase DHAR2 (AT1G75270) (Dixon *et al.* 2002).

Biological significance of conserved modules between environments

In addition to identifying environmentally unique correlated gene expression, the consensus WGCNA is also useful for identifying gene expression modules that are conserved between environments. Functional annotation clustering of genes in each of the 10 consensus modules identified biological processes that can be considered to be robust to environmental variation.

The largest consensus module was con-turquoise, which contained 548 genes. Functional annotation clustering identified numerous clusters, including those related to protein localization, vesicle composition/trafficking, Golgi transport, protein catabolism, and polysaccharide metabolism. The con-red module was found to be enriched for genes related to cell-wall organization and cellulose biosynthesis. The con-brown and con-yellow modules were grouped together in both the H04 and the K05 eigengene clustering (Figure 4, A and B). Taken together, the genes from these modules were functionally annotated as being involved in photosystem repair, plastid localization, and phenylpropanoid biosynthesis. Using all genes contained in the consensus network for GO enrichment analysis, we identified several significantly enriched (corrected $P < 0.01$) biological processes, including glucan metabolic processes, cellulose biosynthesis, cellular protein localization, and plant-type cell-wall biogenesis.

Discussion

eQTL mapping and distribution

In this study, the mapping of eQTL across the wheat genome identified genomic loci that contributed to the genetic regulation of gene expression for many genes, so-called eQTL hotspots. Frequently, these eQTL hotspots were found at the ends of the linkage groups. This pattern may be due to the greater gene density at the telomeric ends of wheat chromosomes (Qi *et al.* 2004). Additionally, eQTL with true peaks at a genomic locus beyond the linkage map coverage would accumulate at the last marker in the linkage group. In barley, eQTL hotspots were found in low-recombination

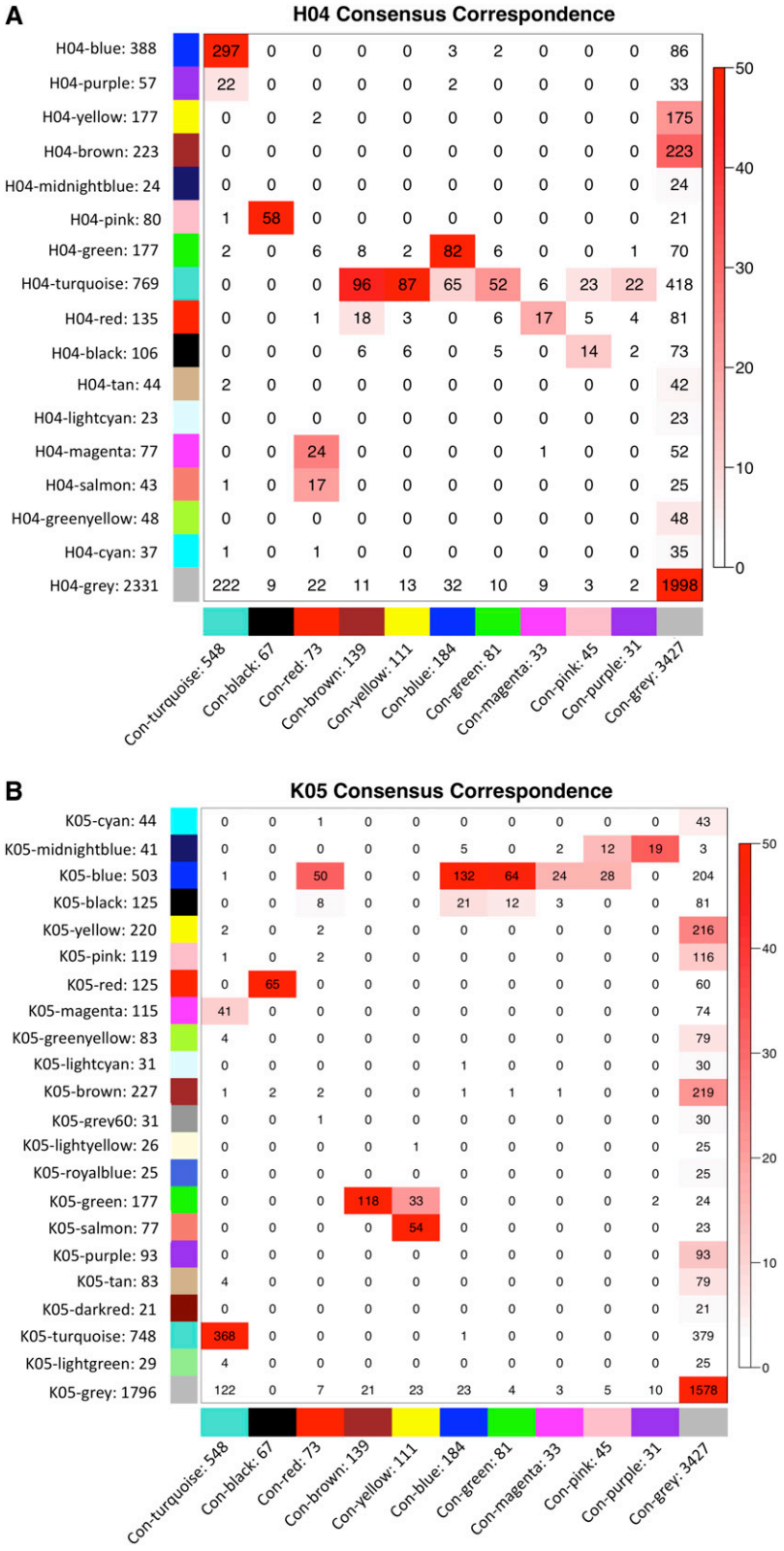

Figure 5 Correspondence of single-environment WGCNA module genes to consensus WGCNA modules for (A) the H04 environment and (B) the K05 environment. Single-environment modules are denoted on the vertical axis with the total number of genes in each module listed next to the label. Consensus modules are shown on the horizontal axis with the total number of genes in each module listed next to the label. The gray module contains all genes not connected in the network. The internal cells correspond to the number of genes common between each individual and consensus module. Shading denotes the significance of the correspondence based on Fisher's exact test with darker shading representing higher significance with a scale of $\log_{10}(1/P\text{-value})$.

centromeric regions and were attributed to the greater physical distance as related to genetic distance (Potokina *et al.* 2008). A better understanding of the eQTL patterns observed in this and other studies will be possible once a fully assembled wheat genome sequence becomes available.

Estimating *cis*- or *trans*-regulation in wheat is hampered by the lack of a fully assembled genome sequence or a high-density SNP map. We used information regarding the position of deletion bin-mapped ESTs to roughly ascertain the mode of regulation. These calculations likely

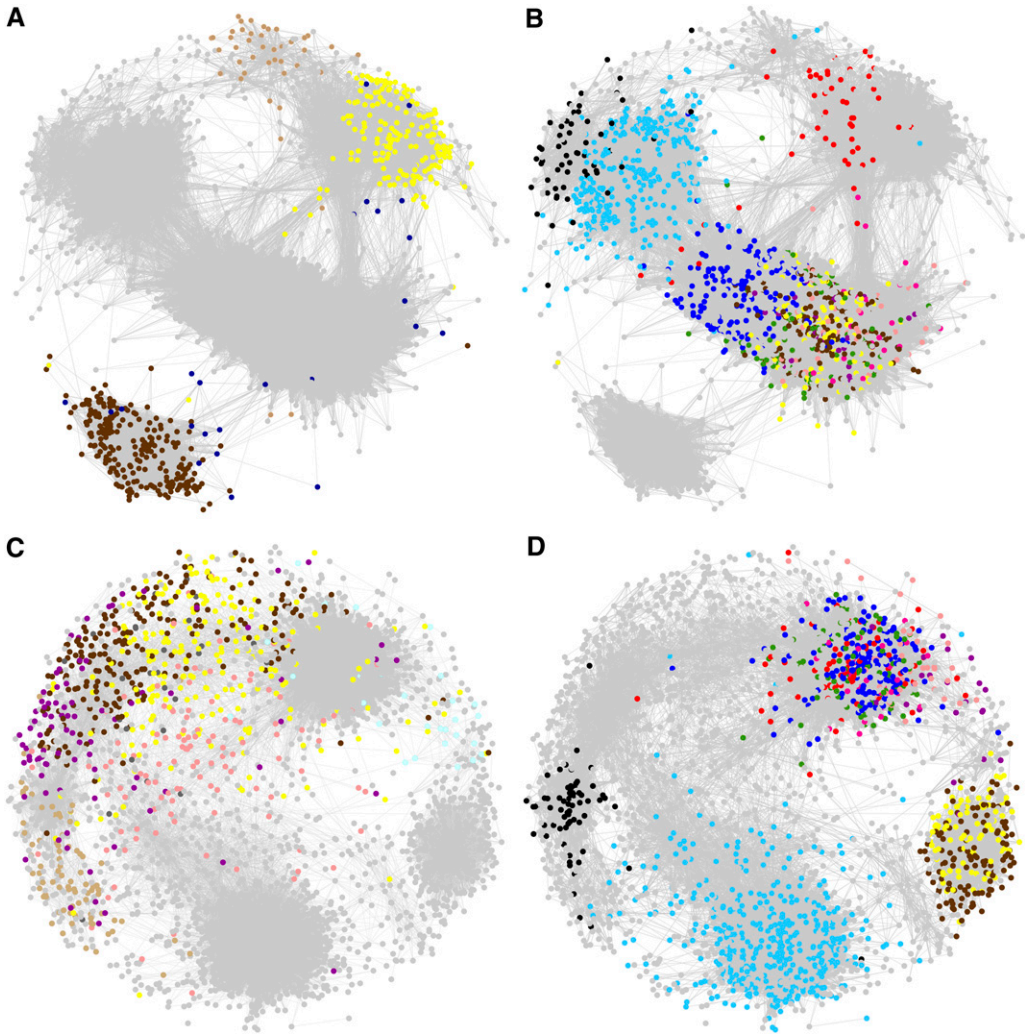

Figure 6 Network view of the H04 and K05 WGCNA created by Cytoscape using an unweighted, spring-embedded layout. (A) The H04 network with nodes colored according to unique individual environment modules. (B) The H04 network colored according to consensus modules. (C) The K05 network colored by unique environment modules. (D) The K05 network colored by consensus modules.

inflate the number of *cis*-eQTL due to the inability to accurately position the mapped ESTs in the linkage map. Even with the probable overestimation, only ~15% *cis*-regulation was detected, which is considerably lower than for other species (Potokina *et al.* 2008). Despite the low levels of *cis*-regulation, the general trend for the eQTL with the highest LOD scores to be due to *cis*-acting factors observed in other studies (Gibson and Weir 2005) was also found in the current study.

The eQTL positions from each environment were compared to identify conserved and unconserved eQTL locations (Figure 2). Along almost the entire linkage map eQTL were identified at shared loci between the H04 and K05 environments. Perhaps more interestingly, a number of eQTL clusters were identified that mapped to different loci in each of the environments. This may be due to transcriptional control of these genes by multiple shared regulators, which contribute more or less to the transcriptional abundance of these genes based on environmental conditions. The reprogramming of transcriptional regulation has been previously observed in barley when comparisons of eQTL positions were made between data from plants challenged with a control or stem rust pathogen (Moscou *et al.* 2011).

Comparison of gene expression networks across environments

The WGCNA method used in this study is a powerful tool for identifying concerted patterns of gene expression, or modules, that can be interrogated for meaningful biological function. Consensus and differential WGCNA methods extend this analysis to two or more datasets and allow for the identification of robust unique and shared gene expression patterns at a systems rather than a gene level (Langfelder and Horvath 2007).

By taking advantage of consensus WGCNA methods, we were able to identify environmentally unique and shared coordinated gene transcription. Functional analysis of the genes for the individual modules identified coordinated regulation related to biological processes that are responsive to environmental conditions. In the H04 environment it appears that co-expression of genes involved in defense responses was unique as compared to the K05 location. Based on the climatological data from the two environments

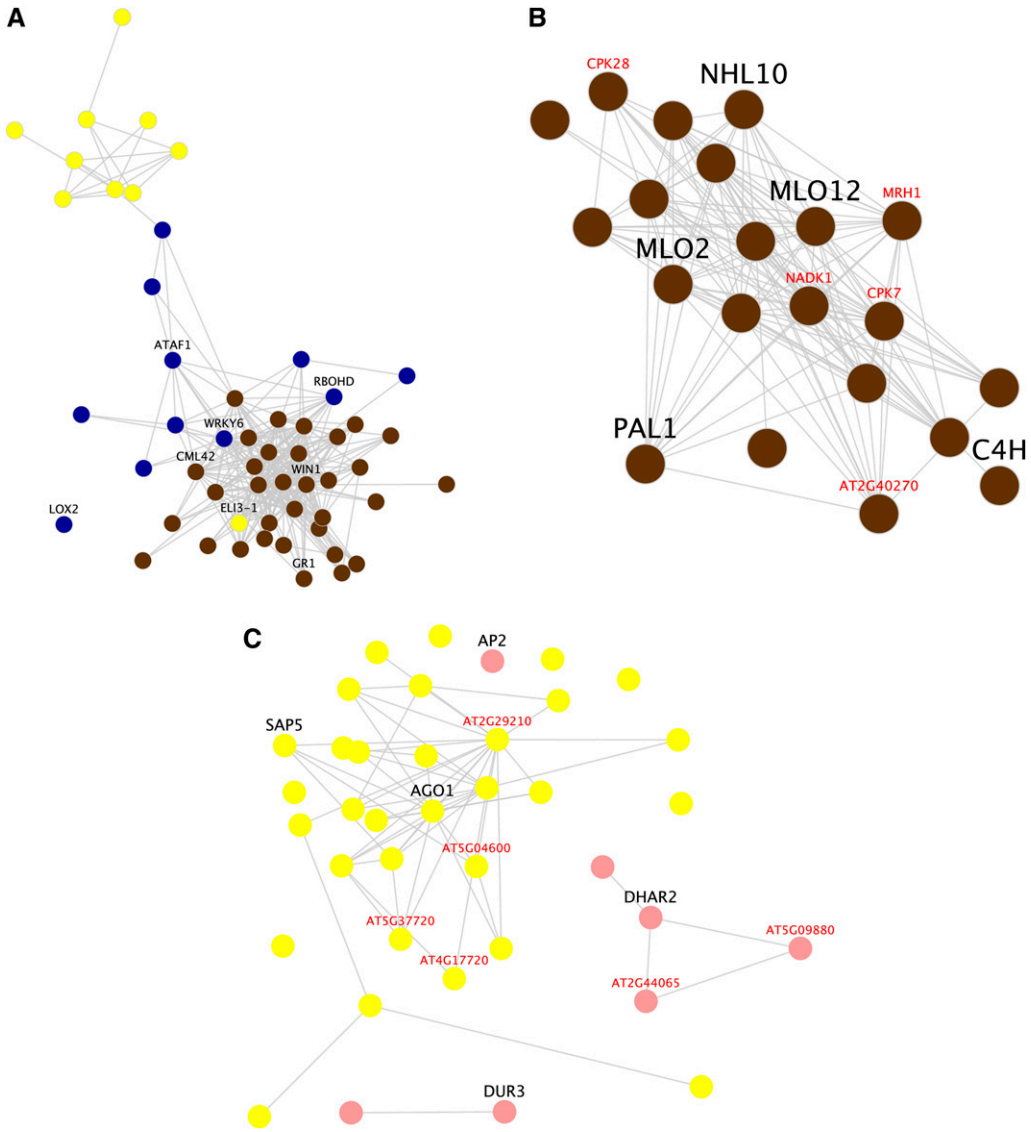

Figure 7 Genetics and environment interact to control biologically meaningful coordinated gene expression. (A) Interaction of genes in H04-specific modules controlled by a locus on chromosome 6D. Genes are colored based on their H04-specific WGCNA module. Labels correspond to homologous *Arabidopsis* genes that function in biotic stress response. (B) Interaction of H04-brown modules genes controlled by a locus on chromosome 3D. Black labels indicate genes homologous to *Arabidopsis* genes associated with biotic stress response. Red labels indicate uncharacterized protein kinases. (C) Interaction network of K05-yellow and K05-pink genes controlled by a locus on chromosome 1B. Known *Arabidopsis* homologs involved in abiotic stress response are labeled in black. Genes with putative roles in RNA-binding are labeled in red.

(Figure S1) one could expect higher pathogen loads and increased disease pressure in the cooler and wetter H04 environment. Conversely, the hotter K05 environment had unique modules related to aspects of heat and other abiotic stresses. Similar types of studies done in conjunction with detailed environmental data would help to connect coordinated gene co-expression with specific environmental variables.

The clustering and functional annotation of the H04-brown, H04-midnight blue, and H04-yellow modules and the coordinated regulation of genes known to play a role in disease resistance suggest that these three modules represent an environmentally specific meta-module involved in the response to pathogens. This meta-module is controlled to a large extent by loci on chromosomes 3D and 6D. Similarly, the K05-yellow and K05-pink modules were not conserved in the consensus network and represent an environment-specific meta-module composed of genes related to RNA processing. A second environment-specific meta-module, composed of the K05-brown, K05-grey60, K05-light cyan, and K05-green/yellow modules was found to be enriched for gene-related heat response. In contrast to the environmentally unique modules, the consensus modules were generally composed of genes related to basic cellular functions, including protein localization and trafficking and genes responsible for structure and maintenance of the photosystem.

One surprising finding is that genes contained in the consensus network had very few eQTL and not one gene had a shared eQTL position across the environments. One might expect highly reproducible genetic control of coordinated gene regulation that is stable in different environments. This makes logical sense, but if one takes into consideration the transcript variation and biological function of these genes, a different conclusion arises. The ability to map eQTL for particular genes requires that genetic variation within the population controls the changes in gene expression. This requirement is unnecessary for genes to be connected in a co-expression network because coordinated changes in expression may be the result of microclimate and other fluctuations

that interact weakly with genetic variation. Connection within a network simply requires that the gene expression changes were sufficiently correlated with changes in other genes.

The observation that genes in consensus modules tend to be important for general biological processes but lack coordinated genetic control suggests that the variation in transcriptional regulators of those genes is more constrained. It has been previously shown that gene essentiality and network connectivity in yeast is correlated with sequence conservation (Carlson *et al.* 2006). Another study in yeast also found that dispensable genes showed higher rates of G×E interaction but found no evidence for a relationship between sequence evolution and G×E interaction (Landry *et al.* 2006). Analyses to test if genetic variation is constrained for consensus regulators await resequencing studies in wheat that would allow for the comparison of genetic variation between genes with environmentally unique or conserved expression patterns.

The study presented here represents a step forward in bringing systems approaches to the highly complex wheat genome. The use of consensus network approaches to compare growing environments uses an established approach in a novel manner. This technique can be applied to any multi-environment gene expression study in plants to help understand environmentally controlled gene expression patterns. Additionally, this study provides a gene co-expression framework that can be used in conjunction with specific seed traits in wheat to help understand the underlying gene expression patterns.

Acknowledgments

The authors thank James Tanaka, David Benscher, Emily Combs, and Roxanne VanWormer for technical assistance in the field and lab. This research was supported by National Research Initiative Competitive grants 2005-35301-15728 and 2006-55606-16629 from the U.S. Department of Agriculture Cooperative State Research, Education, and Extension Service and Hatch project 149-419.

Literature Cited

Bassel, G. W., H. Lan, E. Glaab, D. J. Gibbs, T. Gerjets *et al.*, 2011 Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. Proc. Natl. Acad. Sci. USA 108: 9709–9714.

Basten, C., B. Weir, and Z. Zeng, 1994 Zmap: a QTL cartographer, pp. 65–66 in *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production: Computing Strategies and Software*, edited by C. Smith, J. Gavora, B. Benkel, J. Chesnais, W. Fairfull *et al.* Organizing Committee, 5th World Congress on Genetics Applied to Livestock Production, Guelph, Ontario, Canada.

Basten, C., B. Weir, and Z.-B. Zeng, 2002 *QTL Cartographer*. Department of Statistics, North Carolina State University, Raleigh, NC.

Baumberger, N., and D. C. Baulcombe, 2005 Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc. Natl. Acad. Sci. USA 102: 11928–11933.

Bowtell, D., and J. Sambrook, 2003 *DNA Microarrays: A Molecular Cloning Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Brem, R. B., and L. Kruglyak, 2005 The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc. Natl. Acad. Sci. USA 102: 1572–1577.

Brem, R. B., G. Yvert, R. Clinton, and L. Kruglyak, 2002 Genetic dissection of transcriptional regulation in budding yeast. Science 296: 752–755.

Carlson, M. R., B. Zhang, Z. Fang, P. S. Mischel, S. Horvath *et al.*, 2006 Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. BMC Genomics 7: 40.

Consonni, C., P. Bednarek, M. Humphry, F. Francocci, S. Ferrari *et al.*, 2010 Tryptophan- derived metabolites are required for antifungal defense in the Arabidopsis mlo2 mutant. Plant Physiol. 152: 1544–1561.

Decook, R., S. Lall, D. Nettleton, and S. H. Howell, 2006 Genetic regulation of gene expression during shoot development in Arabidopsis. Genetics 172: 1155–1164.

Dixon, D. P., B. G. Davis, and R. Edwards, 2002 Functional divergence in the glutathione transferase superfamily in plants. Identification of two classes with putative functions in redox homeostasis in Arabidopsis thaliana. J. Biol. Chem. 277: 30859–30869.

Drost, D. R., C. I. Benedict, A. Berg, E. Novaes, C. R. Novaes *et al.*, 2010 Diversification in the genetic architecture of gene expression and transcriptional networks in organ differentiation of Populus. Proc. Natl. Acad. Sci. USA 107: 8492–8497.

Druka, A., E. Potokina, Z. Luo, N. Jiang, X. Chen *et al.*, 2010 Expression quantitative trait loci analysis in plants. Plant Biotechnol. J. 8: 10–27.

Fuller, T. F., A. Ghazalpour, J. E. Aten, T. A. Drake, A. J. Lusis *et al.*, 2007 Weighted gene coexpression network analysis strategies applied to mouse weight. Mamm. Genome 18: 463–472.

Gibson, G., and B. Weir, 2005 The quantitative genetics of transcription. Trends Genet. 21: 616–623.

Holloway, B., S. Luck, M. Beatty, J. A. Rafalski, and B. Li, 2011 Genome-wide expression quantitative trait loci (eQTL) analysis in maize. BMC Genomics 12: 336.

Huang Da, W., B. T. Sherman, and R. A. Lempicki, 2009a Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37: 1–13.

Huang Da, W., B. T. Sherman, and R. A. Lempicki, 2009b Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4: 44–57.

Jansen, R. C., and J. P. Nap, 2001 Genetical genomics: the added value from segregation. Trends Genet. 17: 388–391.

Jordan, M. C., D. J. Somers, and T. W. Banks, 2007 Identifying regions of the wheat genome controlling seed development by mapping expression quantitative trait loci. Plant Biotechnol. J. 5: 442–453.

Kang, M., M. Fokar, H. Abdelmageed, and R. D. Allen, 2011 Arabidopsis SAP5 functions as a positive regulator of stress responses and exhibits E3 ubiquitin ligase activity. Plant Mol. Biol. 75: 451–466.

Keurentjes, J. J., J. Fu, I. R. Terpstra, J. M. Garcia, G. Van Den Ackerveken *et al.*, 2007 Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc. Natl. Acad. Sci. USA 104: 1708–1713.

Kojima, S., A. Bohner, B. Gassert, L. Yuan, and N. Von Wiren, 2007 AtDUR3 represents the major transporter for high-affinity urea transport across the plasma membrane of nitrogen-deficient Arabidopsis roots. Plant J. 52: 30–40.

Landry, C. R., J. Oh, D. L. Hartl, and D. Cavalieri, 2006 Genome-wide scan reveals that genetic variation for transcriptional plas-

ticity in yeast is biased towards multi-copy and dispensable genes. Gene 366: 343–351.
Langfelder, P., and S. Horvath, 2007 Eigengene networks for studying the relationships between co-expression modules. BMC Syst. Biol. 1: 54.
Langfelder, P., and S. Horvath, 2008 WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9: 559.
Langfelder, P., R. Luo, M. C. Oldham, and S. Horvath, 2011 Is my network module preserved and reproducible? PLOS Comput. Biol. 7: e1001057.
Lee, I., B. Ambaru, P. Thakkar, E. M. Marcotte, and S. Y. Rhee, 2010 Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana. Nat. Biotechnol. 28: 149–156.
Li, Y., O. A. Alvarez, E. W. Gutteling, M. Tijsterman, J. Fu *et al.*, 2006 Mapping determinants of gene expression plasticity by genetical genomics in C. elegans. PLoS Genet 2: e222.
Liu, L. H., U. Ludewig, W. B. Frommer, and N. Von Wiren, 2003 AtDUR3 encodes a new type of high-affinity urea/H+ symporter in Arabidopsis. Plant Cell 15: 790–800.
Lorenzo, O., J. M. Chico, J. J. Sanchez-Serrano, and R. Solano, 2004 JASMONATE- INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in Arabidopsis. Plant Cell 16: 1938–1950.
McGrath, K. C., B. Dombrecht, J. M. Manners, P. M. Schenk, C. I. Edgar *et al.*, 2005 Repressor- and activator-type ethylene response factors functioning in jasmonate signaling and disease resistance identified via a genome-wide screen of Arabidopsis transcription factor gene expression. Plant Physiol. 139: 949–959.
Mishina, T. E., and J. Zeier, 2007 Bacterial non-host resistance: interactions of Arabidopsis with non-adapted Pseudomonas syringae strains. Physiol. Plant. 131: 448–461.
Moscou, M. J., N. Lauter, B. Steffenson, and R. P. Wise, 2011 Quantitative and qualitative stem rust resistance factors in barley are associated with transcriptional suppression of defense regulons. PLoS Genet. 7: e1002208.
Mounet, F., A. Moing, V. Garcia, J. Petit, M. Maucourt *et al.*, 2009 Gene and metabolite regulatory network analysis of early developing fruit tissues highlights new candidate genes for the control of tomato fruit composition and development. Plant Physiol. 149: 1505–1528.
Munkvold, J. D., J. Tanaka, D. Benscher, and M. E. Sorrells, 2009 Mapping quantitative trait loci for preharvest sprouting in white wheat. Theor. Appl. Genet. 119: 1223–1235.
Oldham, M. C., S. Horvath, and D. H. Geschwind, 2006 Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proc. Natl. Acad. Sci. USA 103: 17973–17978.
Potokina, E., A. Druka, Z. Luo, R. Wise, R. Waugh *et al.*, 2008 Gene expression quantitative trait locus analysis of 16 000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. Plant J. 53: 90–101.
Pumphrey, M., J. Bai, D. Laudencia-Chingcuanco, O. Anderson, and B. S. Gill, 2009 Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. Genetics 181: 1147–1157.
Qi, L. L., B. Echalier, S. Chao, G. R. Lazo, G. E. Butler *et al.*, 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.
Quackenbush, J., 2002 Microarray data normalization and transformation. Nat. Genet. 32(Suppl.): 496–501.
Saeed, A. I., V. Sharov, J. White, J. Li, W. Liang *et al.*, 2003 TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34: 374–378.
Schadt, E. E., S. A. Monks, T. A. Drake, A. J. Lusis, N. Che *et al.*, 2003 Genetics of gene expression surveyed in maize, mouse and man. Nature 422: 297–302.
Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang *et al.*, 2003 Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13: 2498–2504.
Smith, E. N., and L. Kruglyak, 2008 Gene-environment interaction in yeast gene expression. PLoS Biol. 6: e83.
Thines, B., L. Katsir, M. Melotto, Y. Niu, A. Mandaokar *et al.*, 2007 JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. Nature 448: 661–665.
Wang, X., B. M. Basnayake, H. Zhang, G. Li, W. Li *et al.*, 2009 The Arabidopsis ATAF1, a NAC transcription factor, is a negative regulator of defense responses against necrotrophic fungal and bacterial pathogens. Mol. Plant Microbe Interact. 22: 1227–1238.
West, M. A., K. Kim, D. J. Kliebenstein, H. Van Leeuwen, R. W. Michelmore *et al.*, 2007 Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics 175: 1441–1450.
Xu, X., C. Chen, B. Fan, and Z. Chen, 2006 Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. Plant Cell 18: 1310–1326.
Yang, Y. H., S. Dudoit, P. Luu, D. M. Lin, V. Peng *et al.*, 2002 Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 30: e15.
Zhang, X., A. J. Cal, and J. O. Borevitz, 2011 Genetic architecture of regulatory variation in Arabidopsis thaliana. Genome Res. 21: 725–733.
Zhang, B., and S. Horvath, 2005 A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Molec. Biol. 4: 1128.
Zheng, M. S., H. Takahashi, A. Miyazaki, H. Hamamoto, J. Shah *et al.*, 2004 Up-regulation of Arabidopsis thaliana NHL10 in the hypersensitive response to cucumber mosaic virus infection and in senescing leaves is controlled by signalling pathways that differ in salicylate involvement. Planta 218: 740–750.
Zheng, Z., S. A. Qamar, Z. Chen, and T. Mengiste, 2006 Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. Plant J. 48: 592–605.

Communicating editor: L. M. McIntyre

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150052/-/DC1

Systems Genetics of Environmental Response in the Mature Wheat Embryo

Jesse D. Munkvold, Debbie Laudencia-Chingcuanco, and Mark E. Sorrells

DOI: 10.1534/genetics.113.150052

Guide to Supporting Information:

The supporting information associated with this manuscript provide all the necessary data and code to repeat the described analyses as well as complete results referenced in the paper text.

File **S1** contains the molecular marker names and associated genotype scores across the DH population.

File **S2** summarizes the gene expression values for the DH population, which were used for eQTL mapping and WGCNA. Also included are the microarray probe annotations and the resulting WGCNA module assignments.

File **S3** contains the necessary input files and perl script needed to repeat the eQTL detection using QTLCartographer (http://statgen.ncsu.edu/qtlcart).

File **S4** describes the linkage map, relationship to linkage map bins, and the significant eQTL results for the individual probes in the H04 and K05 environments.

The microarray probes the associated probe ID, sequence and estimated wheat deletion bin map location for the corresponding ESTs can be found in File **S5**.

For repeating the WGCNA analyses File **S6** contains the necessary gene expression input data and R code in markdown format. The R markdown files can be used by the R package Knitr to generate html reports of analysis results (http://yihui.name/knitr/).

For repeating the eQTL enrichment analysis the necessary data files and R code in markdown format can be found in File **S7**.

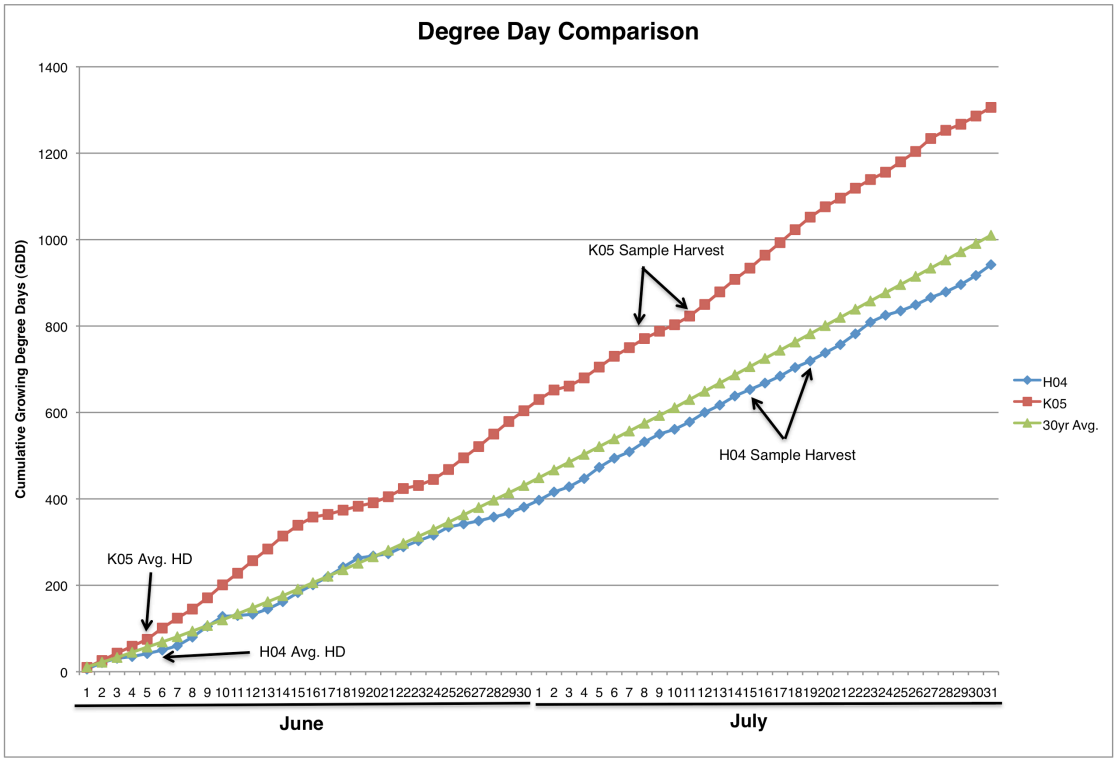

Figure S1 Comparison of cumulative degree growing days between the H04 and K05 environments and the 30 year average. The average heading date (HD) and biological sample harvest dates are labeled for each environment. Climate data from the Northeast Regional Climate Center (http://www.nrcc.cornell.edu/climate/ithaca/).

File S1

Genotype Scores

File S1 contains the genotype scores for the CxC population. These genotype scores and linkage map construction were previously described in Munkvold *et al.* 2009.

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150052/-/DC1.

File S2

Gene Expression Data, Annotation, and Module Assignment

File S2 is a Microsoft Excel document containing four separate worksheets. The first worksheet contains the gene expression data used in eQTL mapping and WGCNA for the H04 environment. The second worksheet contains the gene expression data used in eQTL mapping and WGCNA for the K05 environment. The third worksheet contains the annotation of the microarrays probes based on comparisons with Arabidopsis. The fourth worksheet contains the WGCNA module assignment for the individual environment networks and the consensus network.

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150052/-/DC1.

File S3

Genotype and Linkage Map Input Files for QTLCartographer

File S3 is zip file containing nine different files. These are the input files can be used as input files for QTL mapping in the software program QTLCartographer (http://statgen.ncsu.edu/qtlcart). There are four similar files for each environment used in this study. The .cro files contain the genotype and gene expression phenotype scores and the .map files contain the necessary map information. The files ending in *1 contain the gene expression phenotypes for the first half of the probesets and those ending in 2 contain the gene expression phenotypes for the second half. It was necessary to split the data because together the number of traits exceed the upper limit of the QTLCartographer program.* The file_runpermutes.pl contains the PERL script used to run the QTL permutation testing.

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150052/-/DC1.

File S4

eQTL Results for the H04 and K05 Environments

File S4 is a Microsoft Excel workbook containing the CIM eQTL results for the H04 and K05 environments. The file contains five different worksheets including cross-referencing between linkage group and chromosome, marker names and linkage map positions, cross-referencing between the linkage map and bins, and the CIM eQT results with corresponding permutation thresholds for each environment.

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150052/-/DC1.

File S5

Oligo Sequence Data and Wheat Chromosome Deletion Bin Chromosome Location

File S5 is a Microsoft Excel Workbook containing two worksheets. The first sheet gives the microarray probe oligo identifier and corresponding probe sequence. The second sheet details the wheat chromosome location for each probe as determined by BLAST against the deletion bin-mapped wheat ESTs.

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150052/-/DC1.

File S6

WGCNA Input and R Code

File S6 is a zip file containing five different folders. Each folder contains the input data and an R markdown file necessary to run WGCNA on the wheat expression data. An automated html report of the analysis can be generated using the R markdown and the knitr R package (http://cran.r-project.org/web/packages/knitr/index.html), which is also nicely integrated into the R coding software R Studio (http://www.rstudio.com). The files include the full and trimmed expression data analysis for both the H04 and K05 environments and the WGCNA consensus analysis of the trimmed data from both environments.

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150052/-/DC1.

File S7

eQTL Enrichment Analysis Input and R Code

File S7 is a zip file containing the input files, result files, and R markdown files necessary to test for eQTL enrichment using the Fisher Exact Test. For the H04 and K05 environments the results files summing up the eQTL counts by module and bin and the corresponding p-values are provided. An automated html report of the analysis can be generated using the R markdown and the knitr R package (http://cran.r-project.org/web/packages/knitr/index.html), which is also nicely integrated into the R coding software R Studio (http://www.rstudio.com).

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150052/-/DC1.