2014

eQTL Networks Reveal Complex Genetic Architecture in the Immature Soybean Seed

Yung-Tsi Bolon
University of Minnesota, hsie0024@umn.edu

D. Hyten
Southern Illinois University, david.hyten@unl.edu

James H. Orf
University of Minnesota

Carroll P. Vance
University of Minnesota

Gary J. Muehlbauer
University of Minnesota, muehl003@umn.edu

Published in *The Plant Genome* 7:1 (March 2014), pp. 1–14; doi: 10.3835/plantgenome2013.08.0027

Submitted August 9, 2013.

Supporting information for this article is available following the references and in separate spreadsheet files.

eQTL Networks Reveal Complex Genetic Architecture in the Immature Soybean Seed

Yung-Tsi Bolon,[1] David L. Hyten,[2] James H. Orf,[1] Carroll P. Vance,[1] and Gary J. Muehlbauer[1,3]

1. Department of Agronomy and Plant Genetics, University of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN, 55108
2. Pioneer Hi-Bred International Inc., Johnston, Iowa 50131
3. Department of Plant Biology, University of Minnesota, St. Paul, MN 55108

Corresponding author – Yung-Tsi Bolon, email hsie0024@umn.edu

Abstract

The complex network of regulatory factors and interactions involved in transcriptional regulation within the seed is not well understood. To evaluate gene expression regulation in the immature seed, we utilized a genetical genomics approach on a soybean [*Glycine max* (L.) Merr.] recombinant inbred line (RIL) population and produced a genome-wide expression quantitative trait loci (eQTL) dataset. The validity of the dataset was confirmed by mapping the eQTL hotspot for flavonoid biosynthesis-related genes to a region containing repeats of chalcone synthase (CHS) genes known to correspond to the soybean inhibitor locus that regulates seed color. We then identified eQTL for genes with seed-specific expression and discovered striking eQTL hotspots at distinct genomic intervals on chromosomes (Chr) 20, 7, and 13. The main eQTL hotspot for transcriptional regulation of fatty acid biosynthesis genes also coincided with regulation of oleosin genes. Transcriptional upregulation of genesets from eQTL with opposite allelic effects were also found. Gene-eQTL networks were constructed and candidate regulatory genes were identified from these three key loci specific to seed expression and enriched in genes involved in seed oil accumulation. Our data provides new insight into the complex nature of gene networks in the immature soybean seed and the genetic architecture that contributes to seed development.

Abbreviations: ABC, ATP-binding cassette; ACP, acyl carrier protein; BME, blue micropylar end; Chr, chromosome label; CHS, chalcone synthase; *CTS*, COMATOSE; eQTL, expression quantitative trait locus; FAH, fatty acid hydroxylase; FDR, false discovery rate; GATA, DNA binding motif; LG, linkage group; LOD, logarithm of odds; M × N, Minsoy × Noir 1; QTL, quantitative trait locus; RIL, recombinant inbred line; RNA-seq, RNA sequencing; SNP, single nucleotide polymorphism.

Regulation of gene transcription is an integral function of biological systems. In the last decade, global gene transcription studies have provided an unprecedented view of the genetic architecture of diverse organisms (Brem et al., 2002; Drost et al., 2010; Morley et al., 2004; Schadt et al., 2003; West et al., 2007). Gene eQTL mapping is a powerful tool for identifying the genetic basis of gene expression variation. Analyses of gene transcript accumulation patterns in segregating populations through a genetical genomics approach can help to dissect complex traits (Bystrykh et al., 2005; Chesler et al., 2005; Grupe et al., 2001; Quigley et al., 2009). In plants, eQTL studies have been applied toward understanding development and biotic and abiotic responses (Chen et al., 2010; Druka et al., 2008; Hammond et al., 2011; Moscou et al., 2011). An eQTL study was recently reported for *Arabidopsis thaliana* (L.) Heynh. seed (Cubillos et al., 2012). However, a genome-wide assessment of transcriptional regulation in the immature legume seed has yet to be elucidated.

Because oilseeds are an important source of oil and protein for both human and animal consumption and industrial use, the factors that control accumulation of these traits during soybean seed development are of considerable interest. Identification of genetic factors that control seed composition have been the focus of widespread quantitative trait locus (QTL) mapping efforts in the past (Bolon et al., 2010; Brummer et al., 1997; Delourme et al., 2006; Diers et al., 1992; Fasoula et al., 2004; Hobbs et al., 2004; Lee et al., 1996; Mansur et al., 1993; Panthee et al., 2005). In common, these analyses point to the involvement of multiple genetic loci and an inverse correlation between seed oil and seed protein content (Brim and Burton, 1979). Similar to QTL for phenotypic traits, the loci involved in regulation of gene expression traits may be mapped as eQTL. The combination of genetic mapping with genome-wide gene transcript accumulation profiles in a segregating population can be used to construct transcriptional networks (Jansen and Nap, 2001; Keurentjes et al., 2007) and to link gene regulation to control of seed composition traits.

As a major oilseed crop with the advantage of a sequenced genome (Schmutz et al., 2010), the immature soybean seed is an excellent platform for analysis of transcriptional regulation through eQTL mapping. In soybean, eQTL analyses have been performed on several sets of genes (Song et al., 2013; Yin et al., 2010, 2011), but a genome-wide assessment of soybean eQTL has not been conducted. A well-established RIL population descended from parent soybean cultivars Minsoy (M) and Noir 1 (N) was chosen for its wide genetic (Hyten et al., 2010) and phenotypic (Lark et al., 1995) diversity. Phenotypic differences between Minsoy and Noir 1 and within the M × N RIL population have been observed for many traits, including height, leaf, and flower characteristics (Lark et al., 1995), as well as seed coat pigmentation and seed composition, including seed oil and protein content (Orf et al., 1999). Coincident genetic locations of eQTL and phenotypic QTL provide the basis for further investigation of molecular mechanisms. Previous studies have shown the utility of combined eQTL and QTL mapping to gain insight into regulatory pathways involved

in determining phenotypic traits (Chen et al., 2012; Mehrabian et al., 2005; Moscou et al., 2011; Schadt et al., 2003; Sonderby et al., 2007).

Using genome-wide eQTL data, regulatory networks may be constructed a posteriori (Wu et al., 2008) or a priori (Keurentjes et al., 2007; Kliebenstein et al., 2006; Wentzell et al., 2008; Wentzell et al., 2007). Both approaches may be used to develop gene regulatory connections and to examine genes involved in metabolic pathways in the immature seed. Genome-wide mapping of eQTL also provides data on the overall genetic architecture of transcriptional control. Analyses of eQTL mapping studies in other organisms and from other plant tissues have provided observations on the complex nature of gene regulatory control that include enriched gene categories, transgressive segregation, and biased patterns of allelic influence (Brem and Kruglyak, 2005; Drost et al., 2010; Li et al., 2013; West et al., 2007).

Numerous QTL for seed traits and candidate regulatory factors for influencing seed development and composition have been identified previously. However, gene regulatory connections among multiple loci across the genome are not well understood. In this study, we mapped eQTL across the genome to elucidate gene regulatory relationships and biological pathway associations specific to the immature soybean seed. Two approaches were taken for network analysis in this study. The eQTL for genes of specific seed functions were examined for genome-wide colocalization patterns. Second, major eQTL hotspots were examined for enrichment in annotated functional categories. Our objectives were to assess genome-wide transcriptional regulation patterns concentrated in the immature seed and to evaluate genetic control of transcription in relationship to seed processes important for seed oil accumulation. We show the utility of the genetical genomics approach for analysis of complex gene regulation in the soybean seed and find evidence for three main hotspots of gene expression regulation for major seed process pathways in this population.

Materials and Methods

Plant Material and Sample Processing

Minsoy PI 27890 and Noir 1 PI 290136 soybean cultivars were propagated in the growth chamber. For evaluation of transcriptional profiles in these cultivars during seed development, immature green seed tissue from early- to mid-maturation stages (Bolon et al., 2010; Severin et al., 2010b) corresponding to 10 mg and 100 mg seed weight was collected, and total RNA was extracted from the whole immature green seed. Seed tissue from three plants were pooled for each replicate, and three replicates were obtained per genotype at each stage. Minsoy, Noir 1, and 93 M × N RILs selected for enrichment of diversity in recombination breakpoints were grown in the greenhouse under an initial photoperiod of 14/10 h (day/night) and thermocycle of 22 and 10°C, and monitored under Minnesota field growing conditions. Soybean seed tissue from whole immature green seed from a single stage corresponding to the 10 mg seed weight (Suppl. Fig. S1) was collected for the eQTL study from a minimum of thirty pods from three different plants. Approximately 500 mg of seed tissue was then ground and pooled for RNA extraction. Total RNA from all samples were isolated according to a modified protocol (Wang and Vodkin, 1994) designed for high quality RNA extraction from the seed of soybeans that produce procyanidins during

seed coat development. Total RNA was purified by RNeasy column (Qiagen, Valencia, CA) and evaluated by Bioanalyzer (Agilent, Germany) chip.

Genetic Map Reconstruction

Genomic DNA samples from the M × N RILs were evaluated for single nucleotide polymorphisms (SNPs) on the Illumina Goldengate platform using Universal SNP Beltsville Agricultural Research Center markers (Hyten et al., 2010). Marker data from 1536 loci was obtained to reconstruct a 2500 cM soybean genetic map from 557 SNP markers found to segregate within the population. The SNP markers were subjected to a basic local alignment search tool (Altschul et al., 1990) analysis to obtain the physical SNP marker alignment position relative to the 'Williams82' reference soybean genome sequence (Schmutz et al., 2010). A soybean genetic map was reconstructed through JoinMap, and 24 linkage groups (LGs) were assembled. The resulting 2500 cM map possessed an average intermarker distance of 4.74 cM across the ~1100 Mb soybean genome sequence.

Microarray Processing and Data Analyses

RNA sample processing, labeling, and hybridization to the Soy Genome Affymetrix GeneChip were performed as described (Bolon et al., 2010). Gene expression profiles on over 37,500 probesets were analyzed as described (Bolon et al., 2010) for Minsoy and Noir 1 parents and 93 members of the M × N RIL population. Datasets were deposited under Experiment GM48 in the Plant Expression Database (Dash et al., 2012). Differentially accumulated gene transcripts between Minsoy and Noir 1 were detected using Student's *t* test with a 5% false discovery rate (FDR) threshold (Storey and Tibshirani, 2003). Transgressive segregation was assessed for each probeset based on the number of segregants with transcript accumulation levels of at least one standard deviation beyond the parental means (Brem and Kruglyak, 2005; West et al., 2007).

eQTL Mapping and Analyses

Composite interval mapping (Zeng, 1994) was performed using QTL Cartographer v. 1.17 with a walking speed of 2 cM and window size of 10 cM for each trait using the genetic map constructed from the M × N RIL population. To declare statistically significant eQTL as described (West et al., 2007), a global threshold was determined based on 100,000 maximum likelihood ratio test statistics from 100 randomly selected gene expression traits using 1000 permutations (Churchill and Doerge, 1994). The global permutation threshold was 15.5698, calculated as the average 95th percentile value of the representative null distribution of the maximum likelihood ratio test statistic. Each eQTL mapping profile was sorted such that the maximum logarithm of odds (LOD) peak separated by at least 2 LOD difference from an adjacent peak on the same chromosome was counted as a separate eQTL. Hotspots of significant eQTL were then identified by permuting the distribution of eQTL across the genetic map and storing the maximum number of eQTLs uncovered per centimorgan bin. After 1000 permutations of 28,470 eQTL across 1511 bins, a hotspot threshold of 39 eQTL corresponding to the 95th percentile of the stored value was obtained. The number of eQTL hotspots with significant ($P < 0.01$) eQTL excess was calculated after accounting for genes per interval and using the contingency chi-squared test as described

(Potokina et al., 2008). Following the referenced protocol, the ratio of genes to eQTLs for this study was 0.66339:0.33661 and this proportion was used to test the null hypothesis that the number of genes and eQTLs in each interval was consistent. At least seven genes and eQTLs were present in each interval for use of the chi-square test. A total of 54 intervals were considered to be significantly ($P < 0.01$) enriched in eQTL following the contingency chi-squared test. The number of eQTL in these intervals also exceeded the permuted hotspot threshold of 39 eQTL. Networks of genes and eQTL were visualized using Cytoscape v. 2.7.0 (Shannon et al., 2003).

Pathway Enrichment Analysis

Gene models from the reference soybean genome sequence (www.phytozome.net, verified 13 Dec. 2013) (Schmutz et al., 2010) were used for this study. Pathway and enrichment analyses were performed using Gene Ontology (Ashburner et al., 2000) annotation categories for genes at eQTL hotspots. Gene Ontology enrichment analyses were conducted with GOEAST (Zheng and Wang, 2008) using Fisher's exact test (Fisher, 1925) and the multitest FDR adjustment (Benjamini and Yekutieli, 2005). Genes with colocalized eQTL at the hotspot were also represented in metabolic categories using Mapman (Thimm et al., 2004). Genes were represented by color for the allele (red = Minsoy, blue = Noir 1) with additive effect, and by intensity based on LOD score for the eQTL in the diagram.

Phenotypic QTL Mapping

Seed coat pigmentation on seed from the M × N RILs was quantified by ImageJ (NIH). Seed oil and protein was evaluated by near-infrared reflectance spectroscopy or petroleum ether extraction and combustion (Eurofins) and recorded for Minnesota seed harvests in 1992 and 1993 (Mansur et al., 1993), 1998 (Rosemount, MN; Orf et al., 1999), 2001 (Waseca, MN), 2008 (St. Paul, MN), and 2009 (St. Paul, MN). Phenotypic QTL were calculated using the reconstructed genetic map for the M × N RILs and using QTL cartographer for composite interval mapping as described above. Seed QTL positions were calculated based on the newly refined genetic map using seed oil and protein data collected over the span of two decades. The QTL traces for each phenotype in each environment were plotted collectively to show overlap in high-confidence QTL regions detected on at least two occasions at a minimum LOD of 3.5 (Van Ooijen, 1999).

Results

Genetic Map Reconstruction and Seed QTL for M × N RILs

A RIL population derived from soybean parental genotypes Minsoy and Noir 1 was utilized in this study. Previous studies involving this M × N RIL population relied on genetic maps constructed from simple sequence repeat and restriction fragment length polymorphism marker data (Lark et al., 1995; Mansur et al., 1993). To increase the accuracy of this mapping study, we obtained SNP marker data from 1536 loci across the sequenced soybean genome (Schmutz et al., 2010) and reconstructed a 2500 cM soybean genetic map from 557 markers (Suppl. Table S1) that were found to segregate within the M × N RIL population. Twenty-four LGs with an average intermarker distance of 4.7 cM (Suppl. Table S2)

formed the genetic map for this population, using the soybean genome sequence (Schmutz et al., 2010) for alignment of reference markers.

Multiple QTL for seed composition have been identified in the M × N RILs (Lark et al., 1994; Mansur et al., 1993; Orf et al., 1999). Soybean plants of the Minsoy genotype typically produce yellow seed of higher oil content and lower protein content than the black-seeded Noir 1 genotype. In field trials from 1992 to 2002, the average seed oil content was 17.43 ± 1.03% from Minsoy and 15.29 ± 0.34% from Noir 1. For these same trials, the average seed protein content was 35.16 ± 0.15% from Minsoy and 37.35 ± 0.84% from Noir 1. Within the M × N RILs, the average seed oil content ranged from 13.64 to 19.20%. The average seed protein content in the M × N RILs ranged from 31.14 to 38.24%. We calculated QTL positions based on the newly refined genetic map using seed oil and protein data collected over the span of two decades (Suppl. Table S3).

The combined seed QTL analyses for the M × N RILs highlighted one QTL for seed oil on Chr 8 and two QTL for seed protein on Chr 4 and 6 that were calculated to explain 10 to 38% of observed trait variation in the population. The existence of coincident seed oil and protein QTL with opposite allelic effects (Suppl. Table S3) for the QTL on Chr 6 and 8 also confirm the inverse correlation that has been observed between these traits (Brim and Burton, 1979). Based on the consensus genetic map (soybase.org, verified 13 Dec. 2013) and corresponding SNP marker positions, all three QTL appear to coincide with prior evidence for a seed oil or protein QTL in soybean. The seed protein QTL interval on Chr 4 coincides with Prot 19-1 (Stombaugh et al., 2004) and Prot 7-2 (Orf et al., 1999) seed protein QTL. Likewise, the seed protein and oil QTL on Chr 6 colocalizes with Oil 23-1 (Hyten et al., 2004), Oil 4-6, and Protein 21-3 (Kabelka et al., 2004) seed QTL. The seed oil and protein QTL on Chr 8 coincides with Oil 1-1 (Mansur et al., 1993) and Prot 17–4 (Tajuddin et al., 2003) seed QTL.

Genome-Wide Gene Transcript Accumulation during Early Seed Maturation

As an initial assessment of gene expression trait variation, Minsoy and Noir 1 were evaluated for transcript accumulation profiles of 30,681 genes in early- and mid-maturation stages of the developing soybean seed using total RNA processed for hybridization to the Affymetrix Soy GeneChip. A total of 200 genes (Suppl. Table S4) were found to be significantly differentially expressed between Minsoy and Noir 1 at an FDR of 5% and at a fold-change ratio of two or more. Transcripts for genes involved in lipid transport function were among those that were differentially expressed between the two genotypes.

Transcript accumulation data was collected from the immature green seed stage (Suppl. Fig. S1, Suppl. Table S5) corresponding to early seed maturation at the onset of reserve accumulation from 93 members of the M × N RIL population. It is of interest to note that transcript accumulation data on over 16,344 genes supported nonadditive genetic variation through transgressive segregation where gene transcript accumulation values for the RIL population extended beyond the range of the parental values. Previous studies reported similar phenomenon in other species (Brem and Kruglyak, 2005; Hammond et al., 2011; Potokina et al., 2008; West et al., 2007). Within the M × N RIL population, transgressive segregation described the transcript accumulation patterns of more than 50% of gene expression traits evaluated within the population. Transgressive variation was previously

observed for reproductive, morphological, and seed traits, including seed yield, evaluated within the M × N RIL population (Mansur et al., 1993). These results were in marked contrast to the number of gene transcripts (200) that were found to be differentially expressed between the parents Minsoy and Noir 1.

Detection of Global eQTLs in the Immature Soybean Seed

Composite interval mapping identified 28,470 eQTL for 15,568 unique genes expressed in the immature seed (Suppl. Fig. S1) of the M × N RIL population (Fig. 1A). Zero to six eQTL were mapped for each gene expression trait, and the LOD scores for these eQTL ranged from 3.38 to 89.37. These eQTL were found to explain anywhere from a few percent to almost 100% of the variation seen in a given gene expression trait; however, the range of attributed variation was found to lie predominantly between 10 and 20%. A small portion of the mapped eQTL (3824 or 13.4%) was categorized as *cis*-acting regulators based on the proximity of the physical gene location to the eQTL genetic position, with the remainder of the eQTL categorized as *trans*-acting regulators. Because these eQTL have yet to be demonstrated to act in *cis* or *trans*, they may be more appropriately termed local or distant regulators (Rockman and Kruglyak, 2006), respectively, but the more familiar terms *cis*-acting and *trans*-acting are used for clarity. In this study, the eQTL of a gene was defined as *cis*-acting if the gene mapped to the same chromosome and was located within 1.575 Mb of the physical location of the SNP marker near the eQTL. This distance was based on the average intermarker spacing of the SNP markers used for the genetic map. Increasing the allowed distance to 5 Mb or 10 Mb extended the number of candidate *cis*-acting eQTL to approximately 19 or 21% of the total number of eQTL. The proportion of eQTL on any given chromosome that was found to be *cis*-acting varied from 10.5%, on Chr 7, to 49.6%, on Chr 3.

Patterns of Gene Regulation in the Immature Soybean Seed

A number of genetic loci were observed to contain significantly more eQTL than expected to appear by chance distribution (Fig. 1B). A threshold was calculated based on the 95th percentile of the maximum number of eQTL detected at any given locus when 28,470 eQTL were randomly distributed across each interval. Positions where the number of mapped loci peaked above the threshold of 39 eQTL per genetic locus were identified and represent putative regulatory hotspots of gene transcription. Many of these mapped to adjoining positions that likely represent the same hotspot, and after accounting for the number of genes per interval, 54 hotspots were considered enriched for eQTL.

The physical gene position and genetic position of the genome-wide eQTLs for each gene are depicted on an eQTL scatter plot (Fig. 1A). *Cis*-acting eQTL are represented by the diagonal formed across the eQTL scatter plot. All other points on the scatter plot represent *trans*-acting eQTL. The average LOD of a *cis*-acting eQTL was higher (11.65) than the average LOD of a *trans*-acting eQTL (5.65) overall. This result confirms previous reports that *cis*-acting eQTL are typically of greater effect (higher LOD) than *trans*-eQTL in genome-wide studies (Brem and Kruglyak, 2005; Brem et al., 2002; Drost et al., 2010; Keurentjes et al., 2007; Kirst et al., 2005; Morley et al., 2004; Schadt et al., 2003; Vuylsteke et al., 2005; West et al., 2007).

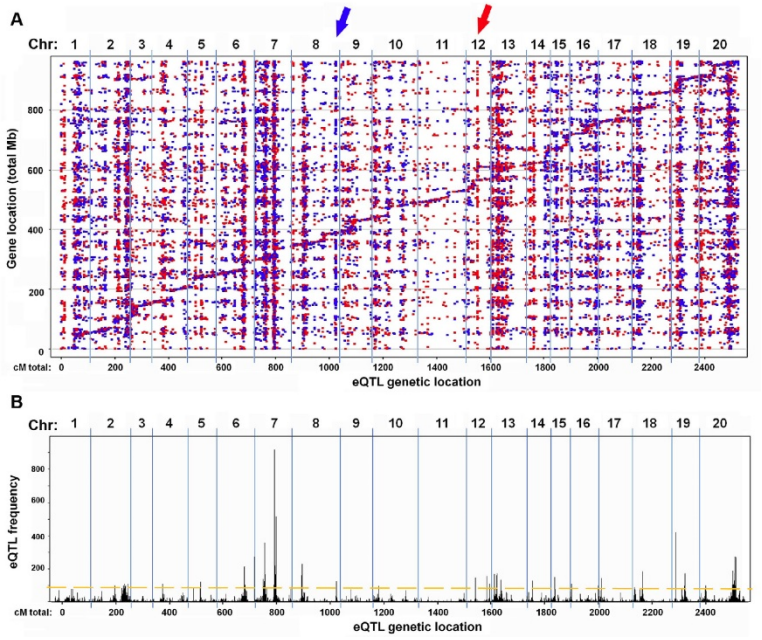


Figure 1. Genome-wide expression quantitative trait loci (eQTL) in the immature seed of the Minsoy × Noir 1 recombinant inbred lines soybean population. (A) An eQTL scatter plot of physical gene location in total megabases versus eQTL genetic location in total centimorgans. Each eQTL point is color-coded to represent transcript accumulation up-regulated by the Minsoy (red) allele or the Noir 1 (blue) allele. Arrows indicate specific examples of allelic bias. (B) The frequency of eQTL mapped to each genetic location is graphed along the 20 soybean chromosomes (Chr). A dashed yellow line denotes the threshold for eQTL hotspots.

Overall, the number of eQTL with allelic effects influencing transcript accumulation in either direction were approximately equal, with 50.18% attributed to the Minsoy allele, and 49.82% to the Noir 1 allele. It was therefore remarkable that a number of eQTL hotspots displayed strong directional bias for allelic effects from a single parent. This directional bias is shown by the color codes for allelic effects in Fig. 1A. On Chr 12, for example, a vertical line of 147 mapped eQTL (red = Minsoy) above genetic position 1554 cM, for example, shows directional bias for the Minsoy allele (136 Minsoy, 11 Noir 1). In the opposite direction, at another eQTL hotspot on Chr 8 at 1023 cM, a directional bias exists for the Noir 1 allele (105 Noir 1, 20 Minsoy). Examples of such directional bias were found on every chromosome.

Validation of the Global eQTL Dataset through eQTL Mapping of Genes Involved in Flavonoid Biosynthesis

Transcriptional regulation of the well-studied flavonoid biosynthesis pathway takes place in the immature seed. Upon mining the eQTL dataset for all genes that annotate to the flavonoid biosynthesis pathway Gene Ontology category, we found that >20% of the eQTL identified for genes in this pathway (adjusted P value = 2.06×10^{-4}) mapped to a Chr 8

interval (~904 total cM, Fig. 2A, Suppl. Table S6). Moreover, the eQTL for flavonoid biosynthesis pathway genes all possessed additive effects with the Noir 1 allele, the genotype with black (versus yellow) seed color. Of these genes, the eQTL for only one gene candidate, the *CHS1* gene (*Glyma08g11610*), was identified as a *cis*-acting regulator. Using a quantitative measurement of seed coat pigmentation and the genetic map assembled for the M × N RIL population, a seed coat pigmentation QTL was also identified over the Chr 8 interval (Suppl. Table S3) and accounted for over 77% of the seed coat pigmentation trait (LOD > 40) (Fig. 2B). The position of this seed coat pigmentation QTL is consistent with the genomic location of a repetitive cluster of CHS genes that controls seed coat pigmentation through generation of small RNAs that downregulate all CHS gene family members (Tuteja et al., 2009).

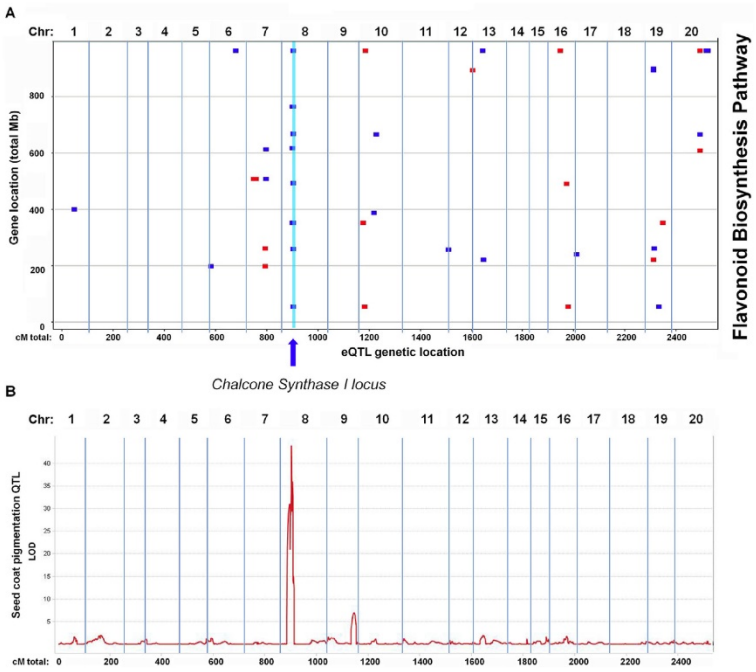


Figure 2. (A) Gene location vs. expression quantitative trait loci (eQTL) genetic location scatter plot of eQTL for genes annotating to the flavonoid biosynthesis pathway. Eight genes involved in flavonoid biosynthesis, particularly chalcone synthase (CHS) genes, mapped to the chromosome (Chr) 8 locus ~904 cM. The Chr 8 *cis*-acting eQTL for *CHS1* was identified at the Inhibitor locus for seed coat color. Each eQTL point is color-coded to represent transcript accumulation upregulated by the Minsoy (red) allele or the Noir 1 (blue) allele. (B) The seed coat pigmentation QTL maps to Chr 8 and colocalizes with the flavonoid biosynthesis eQTL hotspot.

Transcriptional Regulation of Seed-Specific Genes

To identify eQTL hotspots specific to regulation of seed genes, genes that accumulate transcripts in soybean seed tissue alone were identified. The RNA sequencing (RNAseq) data

was obtained from soybean tissues including seed, pod shell, leaf, flower, root, and nodules previously described in a soybean gene expression atlas (Severin et al., 2010b) and combined with unpublished RNA-seq data from a related near-isogenic line (sequence read archive data under BioProject PRJNA208048). Differential gene transcript accumulation from the near-isogenic line pair (HiPro and LoPro) was described for four seed stages (Bolon et al., 2010). HiPro is a high seed protein and low seed oil line that is nearly identical in genotype to the low seed protein and high seed oil line LoPro except for introgression regions (Severin et al., 2010a) that include the major LG I seed protein QTL on Chr 20 (Bolon et al., 2010). Here, the data from all 14 soybean tissues (leaf, flower, pod, shell [2 stages], seed [7 stages], root, and nodule) was utilized for both genotypes to identify genes with seed-specific expression. The eQTL for these seed-specific genes were highlighted from the global eQTL dataset (Suppl. Table S7). Clusters of seed-specific gene eQTL were found at hotspots on Chr 20 (2498 total cM, Fig. 3A) and Chr 13 (1627 total cM, Fig. 3A). The location of these seed-specific eQTL hotspots did not correspond to the eQTL hotspot with the greatest number of eQTL (Chr 7, Fig. 1B) or to the eQTL hotspot for flavonoid biosynthesis (Chr 8, Fig. 2A).

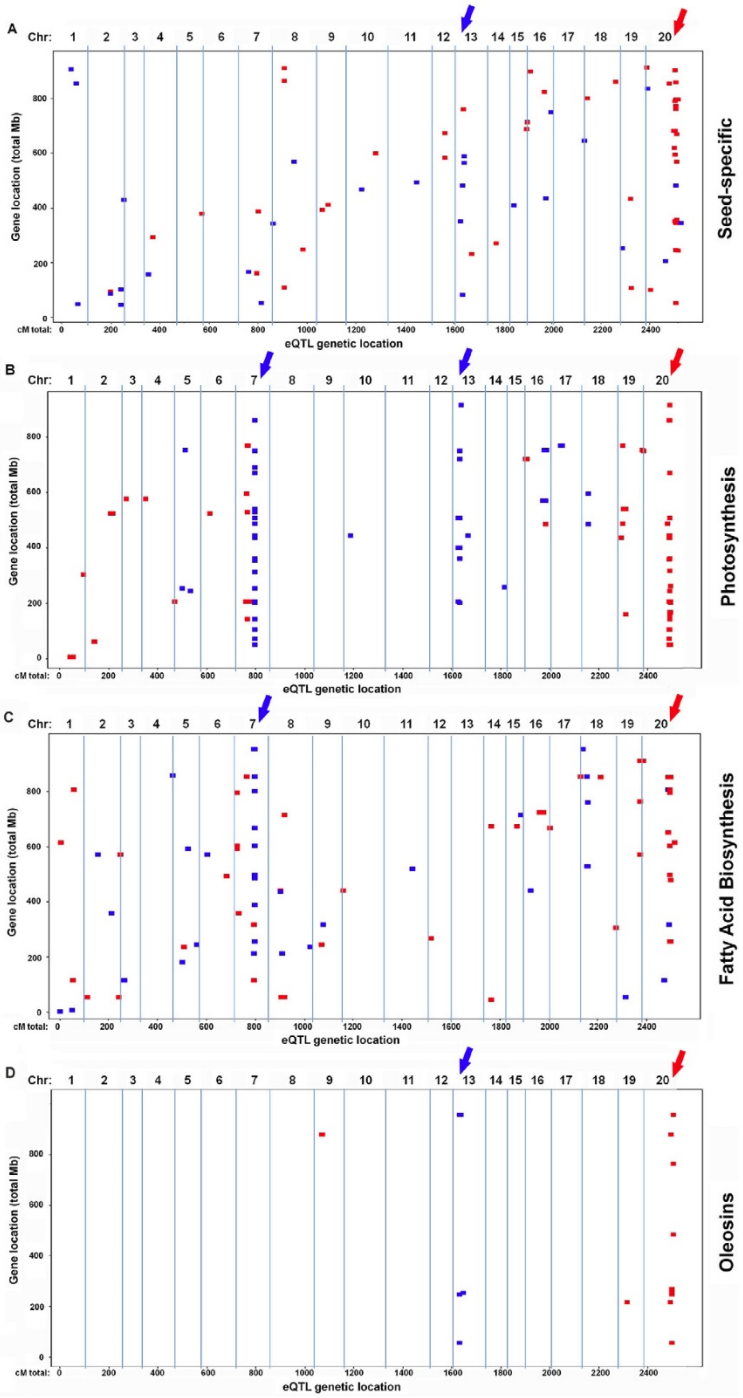


Figure 3. Regulation of seed-specific and seed pathway genes at three expression quantitative trait loci (eQTL) hotspots. Each eQTL point is color-coded to represent transcript accumulation upregulated by the Minsoy (red) or allele or the Noir 1 (blue) allele. Colored arrows highlight the eQTL hotspots indicated. (A) An eQTL scatter plot for seed-specific

genes shows an enrichment of seed-specific eQTL that colocalize to an eQTL hotspot on chromosome (Chr) 20 and another on Chr 13. (B) An eQTL scatter plot for photosynthesis genes reveals eQTL hotspots on Chr 7, 13, and 20. (C) An eQTL scatter plot for genes that annotate to fatty acid biosynthesis show two main loci enriched in fatty acid biosynthesis gene eQTLs on Chr 20 and Chr 7. (D) Mapping of eQTL for oleosin genes shows that the majority of oleosin genes are influenced at one of two loci, Chr 20 or Chr 13, also with directional effects. The eQTL hotspot on Chr 20 is common to all of the above. Transcript accumulation of genes with eQTL in these categories are predominately upregulated with the Noir 1 allele at the Chr 7 and Chr 13 eQTL hotspots whereas they are upregulated with the Minsoy allele at the Chr 20 eQTL hotspot.

Transcriptional Regulation of Specific Seed Functional Pathways

Genes with eQTL at the seed-specific eQTL hotspot (2498 total cM) on Chr 20 were examined for enrichment in specific functional categories. Based on Gene Ontology annotations, the most highly enriched categories at the Chr 20 hotspot were for photosynthesis (adjusted P value = 4.65×10^{-16}) and fatty acid biosynthetic process (adjusted P value = 7.1×10^{-8}) (Suppl. Fig. S2). The eQTL for all genes with either photosynthesis or fatty acid biosynthetic process annotations were subsequently highlighted. Photosynthesis gene eQTL were found to cluster to three regions of the genome, including hotspots on Chr 7, 20, and 13 (Fig. 3B, Suppl. Table S8). Hotspots for eQTL of fatty acid biosynthesis genes were found on Chr 20 and 7 (Fig. 3C, Suppl. Table S9 and S10). Examination of the hotspot on Chr 7 also showed that it was enriched in photosynthesis gene eQTLs (adjusted P value = 3.5×10^{-30}) (Suppl. Fig. S3). It is noteworthy that the majority of seed-specific, photosynthesis, and fatty acid biosynthesis genes with eQTL that mapped to the hotspot on Chr 20 (~2498 cM) showed upregulation of transcript accumulation with the presence of the Minsoy allele (Fig. 3A–3C, Suppl. Fig. S4–S5, Suppl. Tables S7–S9) despite the existence of more eQTL of Noir 1 effect (137 versus 132) at this locus. Surprisingly, eight eQTL for oleosin genes also specifically mapped to this interval on Chr 20 (Fig. 3D, Suppl. Table S9). The expression of these oleosin genes was upregulated with the presence of the Minsoy allele, the genotype with higher seed oil (Orf et al., 1999), consistent with evidence that seeds with higher oil content possess more oleosins (Parthibane et al., 2012a; Siloto et al., 2006).

Multifaceted Regulation of Seed Gene Expression and Relationships with eQTL Hotspots

Complex patterns of directional bias were observed for genes regulated by multiple eQTL. We created a network (Fig. 4, Suppl. Table S11) from the eQTL data to observe the connections among the three eQTL hotspots on Chr 20, 7, and 13 (Fig. 3) with genes for specific seed functional pathways clustered at these hotspots. The majority of these genes with eQTL at the Chr 20 hotspot were found to be upregulated with the Minsoy allele at the Chr 20 hotspot (Fig. 4, Suppl. Table S11). In contrast, the majority of these genes with eQTL at the Chr 7 and 13 hotspots were found to be upregulated with the Noir 1 allele at the respective loci. It is noteworthy that patterns of opposing directional bias were observed for some genes regulated by more than one eQTL. One example of this phenomenon involved a subset of genes with eQTL mapping to the ~2493 to 2498 total cM interval on Chr 20 and

~797 cM on Chr 7 (Fig. 4, Suppl. Fig. S5, Suppl. Table S12). This network showed that transcript levels for a subset of genes are upregulated with the presence of the Minsoy allele at the 2493 locus, and transcript levels of the same genes were upregulated with the presence of the Noir 1 allele at the 797 locus. Genes with upregulated transcript levels with the Minsoy allele included a number of photosynthesis-related genes including genes for plastocyanins, *PsaN* (Photosystem I reaction center subunit PSI-N), *PsaF* (Photosystem I subunit F), and *LHCB3* and *LHCB5* (light-harvesting chlorophyll-binding proteins). Upregulation of transcript levels for fatty acid biosynthesis-related genes, including *FAH1* (Fatty Acid Hydroxylase 1) and *MOD1* (Mosaic Death 1), an enoyl-acyl carrier protein (ACP) reductase subunit of a complex that catalyzes de novo synthesis of fatty acids (Mou et al., 2000) also corresponded to the presence of the Minsoy allele. In addition, all 12 eQTL that mapped to the 2493 to 2498 cM interval for fatty acid biosynthesis gene transcripts were upregulated with the presence of the Minsoy allele, while all nine eQTL that mapped to the 797 cM locus for fatty acid biosynthesis genes were upregulated with the presence of the Noir 1 allele. These findings are consistent with the presence of higher seed oil content in the Minsoy parent versus the Noir 1 parent and the role of photosynthesis and fatty acid biosynthesis in seed oil accumulation.

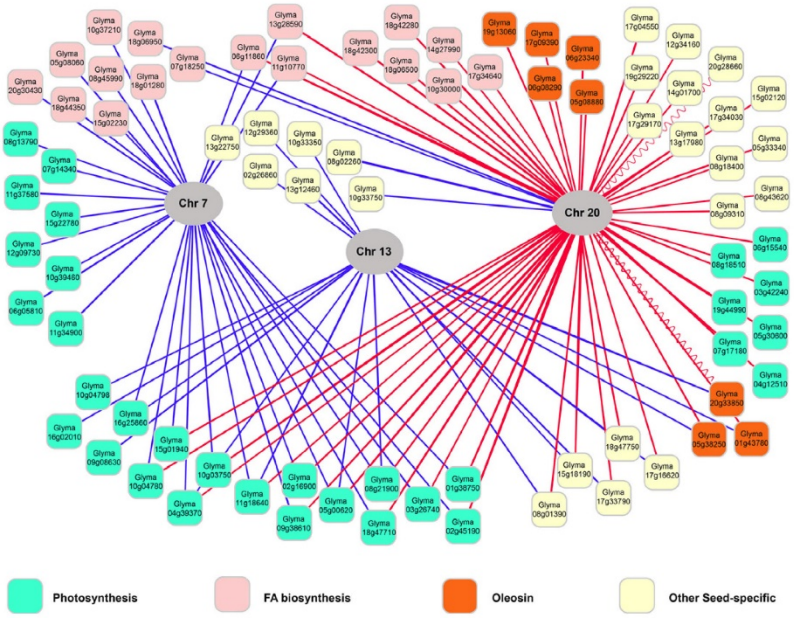


Figure 4. Connections among three major expression quantitative trait loci (eQTL) hotspots involved in major seed and seed-specific processes. A network representation depicts interactions among photosynthesis, fatty acid (FA) biosynthesis, oleosin, and other seed-specific genes with shared eQTL depicted separately in figure 3. Supplemental Table S8 displays the gene and eQTL data represented in this diagram. Gray nodes at the top represent the three major eQTL hotspots at loci on chromosomes (Chr) 7 (~797 total cM), 13 (~1627 total cM), and 20 (2493–2498 total cM). Green nodes = photosynthesis genes. Pink nodes = fatty acid biosynthesis genes. Orange nodes = oleosin genes. Yellow nodes =

seed-specific genes other than the oleosin genes. Connections for genes with transcript accumulation upregulated by the presence of the Minsoy allele are shown with red lines. Connections for genes with transcript accumulation upregulated by the presence of the Noir1 allele are shown with blue lines. Waved lines indicate genes that are within *cis*-acting distances from the hotspot locus.

Regulatory Gene Candidates at the Chromosome 20 Seed eQTL Hotspot

Two eQTL were found to be candidate *cis*-acting regulators out of the 269 eQTL at the 2498 locus (Suppl. Table S9): one of the oleosin genes (*Glyma20g33850*) and a *BME3* (Blue Micropylar End 3) GATA (DNA binding motif) transcription factor gene (*Glyma20g32050*). Although several oleosin genes exist in soybean, the oleosin gene *Glyma20g33850* aligned with the greatest homology to the *OLE3* (Oleosin 3) gene in peanut (*Arachis hypogaea* L.). Oleosin 3 from the immature peanut seed was recently shown to possess diacylglycerol biosynthesis and phosphatidylcholine hydrolysis enzymatic activity that provides evidence for a direct role in increasing oil content through biosynthesis of triacylglycerol from monoacylglycerol (Parthibane et al., 2012a). Sequencing of the *BME3* soybean gene in Minsoy versus Noir 1 genotypes revealed sequence polymorphisms corresponding to missense mutations in four amino acids (T→A, Q→L, V→F, S→T) conserved between soybean and Arabidopsis BME3 transcription factors. The binding motif for *BME3*, WGATAR, was also found in the promoter regions of 251 out of 257 genes with eQTL mapping to the 2498 locus.

Examination of all 319 genes located at the Chr 20 eQTL hotspot showed that there were only eight seed-specific genes and 34 transcription factor genes at this location. Gene transcript accumulation data from RNA-seq profiles compiled from near-isogenic lines HiPro and LoPro with contrasting seed protein and oil content (Bolon et al., 2010; Severin et al., 2010b) show that the seed-specific gene with the highest expression at this locus is the oleosin gene *Glyma20g33850* (Fig. 5A, Suppl. Table S13). Moreover, gene transcript accumulation was highest in the genotype with higher seed oil content (LoPro, Fig. 5B, Suppl. Table S13). Among the transcription factor genes at this locus, transcript accumulation for the *BME3* transcription factor gene *Glyma20g32050* was also the highest, with slightly higher overall transcript accumulation in the LoPro genotype (Fig. 5C, Suppl. Table S13). High levels of gene expression highlighted these two genes from among the 319 genes shown to reside at the Chr 20 eQTL hotspot in the reference soybean genome. Moreover, these same two genes (*Glyma20g33850* and *Glyma20g32050*) were the only genes at this locus with seed eQTL in the M × N RILs that mapped back to the Chr 20 eQTL locus.

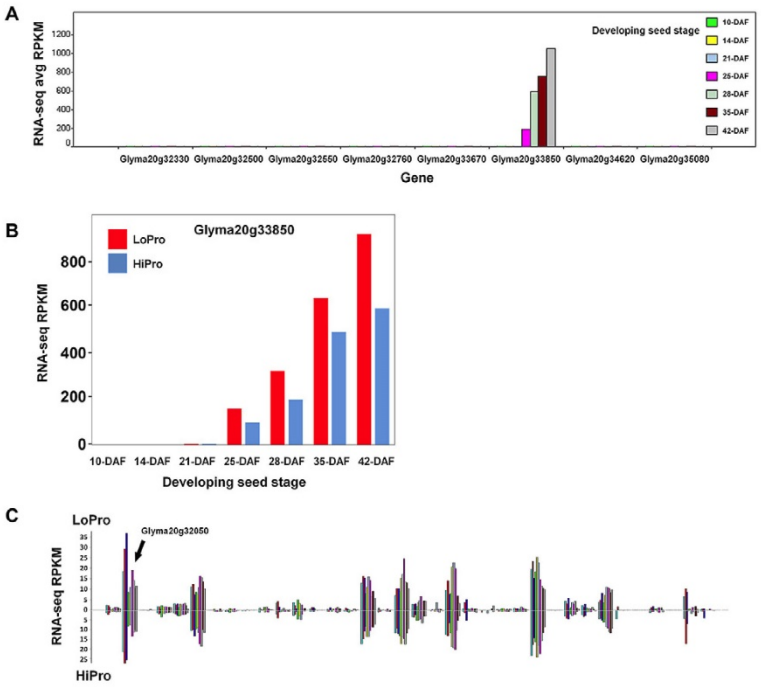


Figure 5. Ribonucleic acid sequencing (RNA-seq) evidence for genes located at the expression quantitative trait loci (eQTL) hotspot on chromosome (Chr) 20. (A) The *y* axis shows RNA-seq RPKM (reads per kilobase of transcript per million mapped reads) counts for each tissue represented as a colored bar. Seed-specific genes at the locus are shown across the *x* axis with colored bars representing different seed stages. Eight seed-specific genes reside at this locus, and the seed-specific gene with the highest evidence of transcript accumulation is the oleosin gene *Glyma20g33850*. (B) The *y* axis shows RNA-seq RPKM counts for each genotype (LoPro = Red, HiPro = Blue) represented as a colored bar. The *x* axis shows a timeline of seed developmental stages. Transcript accumulation of *Glyma20g33850* is higher in LoPro than HiPro during seed development. (C) The *y* axis shows RNA-seq RPKM counts for each tissue represented as a colored bar (leaf, flower, pod, shell [stages –1 and –2], seed [stages –2, –1, 0, 1, 2, 3, 4], root, and nodule tissues). Transcription factor genes at the locus are shown across the *x* axis. The transcription factor gene at this locus with the greatest evidence for transcript accumulation is the GATA (DNA binding motif) gene *Glyma20g32050*. Supplemental Table S13 shows the RNA-seq read evidence for genes in this region in LoPro vs. HiPro across this range of tissues.

eQTL that Colocalize with Seed Phenotypic QTL

To identify other potential genes and pathways that correlate with seed oil and protein accumulation in the immature seed, we also examined eQTL for colocalization with seed oil and protein QTL locations mapped in the M × N RIL population. Expression QTL for 1598 unique genes were found to map to seed oil and protein QTL intervals in this population (Suppl. Fig. S6, purple bars represent seed QTL intervals from Suppl. Table S3). A list of 129 unique genes with *cis*-acting eQTL that colocalized to the region of a seed protein or oil QTL was compiled and included genes with lipid-associated annotations (Suppl.

Table S14). A *cis*-acting eQTL that overlaps an oil QTL on Chr 8 was also mapped for a gene with homology to the *COMATOSE* (*CTS*) ATP-binding cassette (ABC) transporter gene in Arabidopsis.

Discussion

Gene regulation within the immature soybean seed is a multifaceted process that involves many regulatory factors and network connections. Through global eQTL mapping, we discovered striking patterns of directional bias and identified interactions across the genome that provide a rich resource for gene regulatory network construction in soybean. We provide a genomic view of distinct clusters of regulatory action on seed pathways and genes specific to the seed, including processes known to be involved in seed oil accumulation. We show evidence for three major eQTL hotspots important for regulatory control of seed processes through complex gene interactions and identify candidate regulatory genes in the immature seed.

A Positive Control for the Identification of a Transcriptional Regulator through Genome-Wide eQTL Analysis

Results from this eQTL study describe enrichment of eQTL for flavonoid biosynthesis genes within a seed coat pigmentation QTL interval. Among the eQTL at this interval, we identified a candidate for *cis*-acting transcriptional regulation involving the *CHS1* gene on Chr 8. The *CHS1* gene lies within a region of repetitive CHS genes which have been shown to correspond to the soybean Inhibitor locus (Clough et al., 2004; Tuteja et al., 2004). This cluster of CHS genes regulates seed color through production of siRNAs that inhibit expression of CHSs and thus flavonoid synthesis and seed coat pigmentation (Tuteja et al., 2009). This evidence provides a positive control for our eQTL study and demonstrates the capacity to map true regulatory regions and to detect the involvement of specific genes at eQTL hotspots through analysis of this eQTL dataset. From the colocalized eQTL within this interval, our dataset also provides evidence for *trans*-acting regulation of other CHS and flavonoid biosynthesis pathway genes and novel evidence for regulation of additional genes on Chr 6, 13, 15, 17, and 20 (Suppl. Table S3).

Regulatory Gene Candidates at eQTL Hotspots

Confirmation of the localization of the regulatory region for flavonoid biosynthesis genes and seed pigmentation to the CHS repetitive region containing the *CHS1* gene (*Glyma08g11610*) led to the search for genes involved in regulation at other eQTL hotspots. The major eQTL hotspot on Chr 20 was of specific interest for discovery of gene candidates important for regulation of other seed processes, especially those specific to seed tissue. At the highlighted eQTL hotspot on Chr 20, *cis*-acting eQTL for two genes were identified, including an oleosin gene (*Glyma20g33850*) and a *BME3* GATA transcription factor gene (*Glyma20g32050*). These two genes may play a role in the regulation of the important seed processes and seed-specific genes localized to this hotspot. Oleosins provide the structure for accumulation of oil in the seed. Because of the role of oleosin in both sequestering lipid in the seed and recent evidence for oleosin enzymatic activity involved in triacylglycerol

biosynthesis (Parthibane et al., 2012a, 2012b), these genes are important for the storage and accumulation of seed oil. Oleosin gene mutants in Arabidopsis have been found to affect oil body formation and seed germination (Siloto et al., 2006). *BME3* Arabidopsis mutants have also been shown to affect seed germination (Liu et al., 2005). For both genes, the presence of the higher seed oil Minsoy allele corresponded to upregulation of gene expression. Combined with evidence for GATA binding sites in promoter regions of genes with eQTL at this locus, and the presence of sequence polymorphisms in the GATA gene between genotypes, this suggests a possible role for the transcription factor and oleosins in the regulation of seed oil content.

Colocalization of eQTL with phenotypic QTL may be an indicator of the association among regulation of specific gene transcript levels with certain traits. An example of this phenomenon was seen in this study with the colocalization of the flavonoid biosynthesis gene eQTL hotspot and the major seed pigmentation control locus. Because fatty acid biosynthesis, photosynthesis, and oleosins have been shown to play a role in seed oil accumulation and storage (Barker et al., 2007; Huang, 1996; Ruuska et al., 2004; Ruuska et al., 2002), it may be hypothesized that a seed oil QTL would colocalize with the eQTL hotspots for the genes in these pathways. However, the major seed eQTL hotspots enriched in these pathways that were identified in this study did not colocalize directly to seed oil and protein QTL in the M × N RILs population (Suppl. Fig. S6). This phenomenon may be attributed to the use of whole immature seed versus separation of cotyledons and seed coat, and the mapping limitations encountered for complex traits that involve polygenes of varying effects. The mapping of eQTL hotspots that share regulation of fatty acid biosynthesis-related genes with opposing allelic bias may contribute to the complexity. The data were also collected from a single immature seed stage and focused on gene transcript accumulation, a single step at a specific time in the chain of events including posttranscriptional control, protein expression and transport, enzyme activity, pathway progression, and product storage involved in seed oil accumulation. Other factors, such as the lack of sufficient pathway substrates, for instance, may render the effects of increased fatty acid biosynthesis gene transcripts and oleosin gene transcripts ineffectual in changing final seed oil content in certain populations.

Additional Gene Candidates for Control of Seed Oil

A number of gene candidates for the control of seed oil content identified in previous genetic studies were found to correlate to genes and loci identified in this study. The phosphatidic acid phosphatase gene (*At1g15080*) was identified as a gene candidate for control of seed oil in Arabidopsis (Hobbs et al., 2004), and genetic control of transcript accumulation for the soybean phosphatidic acid phosphatase gene homolog (*Glyma10g41580*) colocalized to the Chr 4 seed oil and protein QTL in this study. The ACP 1 gene (*At3g05020*) was found to be upregulated in developing Arabidopsis seeds (Bonaventure and Ohlrogge, 2002) and identified as a gene candidate for control of seed oil (Hobbs et al., 2004). An eQTL for the homologous ACP 1 soybean gene (*Glyma10g30000*) mapped to the fatty acid biosynthesis-enriched gene expression QTL hotspot on Chr 20 described in this study. A long chain acyl-CoA synthetase gene (*At2g47240*) was identified as a gene candidate for control of seed oil in Arabidopsis (Hobbs et al., 2004) and *Brassica napus* L. (Zhao et al.,

2012). An eQTL for a paralogous soybean long chain acyl-CoA synthetase gene (*Glyma10g-01400*) also colocalized to the Chr 8 seed oil QTL identified in this study.

Genes with *cis*-acting eQTL that overlap seed phenotypic QTL were identified. These candidates include a soybean gene homolog (*Glyma08g14480*) of ABC transporter *CTS* (*At4g39850*). The ABC transporter *CTS* regulates transport of acyl-CoAs into the peroxisome (Nyathi et al., 2010). Arabidopsis *CTS* mutants have been found to affect seed germination and lipid metabolism (Footitt et al., 2002) and may affect seed oil content by inhibiting lipid breakdown. It is interesting to note that more than one seed QTL coincided at the Chr 8 interval in this population. A *cis*-acting eQTL for a soybean long-chain acyl-CoA synthetase gene homolog (*Glyma19g40610*) also colocalized to a fatty acid QTL on Chr 19 (data not shown) consistent with the role of long chain acyl-CoA synthetase in fatty acid breakdown.

Coordination of Gene Regulatory Action at Seed eQTL Hotspots

In soybean, a diploidized tetraploid organism that has undergone two rounds of major duplication (Shoemaker et al., 2006), the presence of duplicate genes provides the opportunity to study divergence in gene expression and transcriptional regulatory control (Lin et al., 2011). Gene expression divergence of duplicate genes characterizes and contributes to gene retention and functional evolution. The observation that eQTL regulatory hotspots on two different chromosomes (797 total cM on Chr 7 and 2493 total cM on Chr 20) act on the same network of genes with opposing allelic bias raises the possibility of duplicate gene regulatory action arising from disparate parental genotypes. Although the reference soybean genome sequence (Schmutz et al., 2010) does not show duplicate genes between these particular regions, this is a potential explanation for multiple contributing eQTL hotspots acting on the same set of genes in this population. Another possibility is that a common *trans*-acting factor is encoded by segregating homologous genes in nonallelic locations within the two parental genotypes. Observations of apparently de novo copy number variation in maize RIL populations, for example, have led to a model that involves segregation of single-copy homologous sequences at nonallelic locations in the parental inbred lines (Liu et al., 2012). A transcription factor gene, for instance, could be present in a form that results in expression upregulation of a network of genes in Minsoy and present in a different form at a separate location on another chromosome in Noir 1. Further integration of genetical genomics analyses with studies on gene divergence across species may provide insight into the adaptability of *cis-* and *trans*-acting control and the functional polymorphisms that characterize the method and timing of evolutionary change (Leach et al., 2007).

Acknowledgments – We thank Renee Schirmer for assistance in sampling and processing tissues and Dr. Bruna Bucciarelli for photography. We also acknowledge the use of resources at the Minnesota Supercomputing Institute at the University of Minnesota and thank Dr. Zheng-Jin Tu for his support. This work was supported by the USDA-ARS Current Research Information System (Grant no. 3640-21000-024-00D) and the generosity of the Minnesota Soybean Research and Promotion

Council and the United Soybean Board. Mention of trade names or commercial products in this report is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

References

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:494–498.

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock. 2000. Gene ontology: Tool for the unification of biology. Nat. Genet. 25:25–29. doi:10.1038/75556

Barker, G.C., T.R. Larson, I.A. Graham, J.R. Lynn, and G.J. King. 2007. Novel insights into seed fatty acid synthesis and modification pathways from genetic diversity and quantitative trait loci analysis of the Brassica C genome. Plant Physiol. 144:1827–1842. doi:10.1104/pp.107.096172

Benjamini, Y., and D. Yekutieli. 2005. False discovery rate-adjusted multiple confidence intervals for selected parameters. J. Am. Stat. Assoc. 100:71– 81. doi:10.1198/016214504000001907

Bolon, Y.T., B. Joseph, S.B. Cannon, M.A. Graham, B.W. Diers, A.D. Farmer, G.D. May, G.J. Muehlbauer, J.E. Specht, Z.J. Tu, N. Weeks, W.W. Xu, R.C. Shoemaker, and C.P. Vance. 2010. Complementary genetic and genomic approaches help characterize the Linkage Group I seed protein QTL in soybean. BMC Plant Biol. 10:41–66. doi:10.1186/1471-2229-10-41

Bonaventure, G., and J.B. Ohlrogge. 2002. Differential regulation of mRNA levels of acyl carrier protein isoforms in Arabidopsis. Plant Physiol. 128:223–235. doi:10.1104/pp.010397

Brem, R.B., and L. Kruglyak. 2005. The landscape of genetic complexity across 5700 gene expression traits in yeast. Proc. Natl. Acad. Sci. USA 102:1572–1577. doi:10.1073/pnas.0408709102

Brem, R.B., G. Yvert, R. Clinton, and L. Kruglyak. 2002. Genetic dissection of transcriptional regulation in budding yeast. Science 296:752–755. doi:10.1126/science.1069516

Brim, C.A., and J.W. Burton. 1979. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Sci. 19:494–498. doi:10.2135/cropsci1979.0011183X001900040016x

Brummer, E.C., G.L. Graef, J.H. Orf, J.R. Wilcox, and R.C. Shoemaker. 1997. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37:370–378. doi:10.2135/cropsci1997.0011183X00 3700020011x

Bystrykh, L., E. Weersing, B. Dontje, S. Sutton, M.T. Pletcher, T. Wiltshire, A.I. Su, E. Vellenga, J. Wang, K.F. Manly, L. Lu, E.J. Chesler, R. Alberts, R.C. Jansen, R.W. Williams, M.P. Cooke, and G. de Haan. 2005. Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics.' Nat. Genet. 37:225–232. doi:10.1038/ng1497

Chen, X., W. Guo, B. Liu, Y. Zhang, X. Song, Y. Cheng, L. Zhang, and T. Zhang. 2012. Molecular mechanisms of fiber differential development between G. barbadense and G. hirsutum revealed by genetical genomics. PLoS ONE 7:e30056. doi:10.1371/journal.pone.0030056

Chen, X., C.A. Hackett, R.E. Niks, P.E. Hedley, C. Booth, A. Druka, T.C. Marcel, A. Vels, M. Bayer, I. Milne, J. Morris, L. Ramsay, D. Marshall, L. Cardle, and R. Waugh. 2010. An eQTL analysis of partial resistance to *Puccinia hordei* in barley. PLoS ONE 5:e8598. doi:10.1371/journal.pone .0008598

Chesler, E.J., L. Lu, S. Shou, Y. Qu, J. Gu, J. Wang, H.C. Hsu, J.D. Mountz, N.E. Baldwin, M.A. Langston, D.W. Threadgill, K.F. Manly, and R.W. Williams. 2005. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat. Genet. 37:233–242. doi:10.1038/ng1518

Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963–971.

Clough, S.J., J.H. Tuteja, M. Li, L.F. Marek, R.C. Shoemaker, and L.O. Vodkin. 2004. Features of a 103-kb gene-rich region in soybean include an inverted perfect repeat cluster of CHS genes comprising the I locus. Genome 47:819–831. doi:10.1139/g04-049

Cubillos, F.A., J. Yansouni, H. Khalili, S. Balzergue, S. Elftieh, M.L. Martin-Magniette, Y. Serrand, L. Lepiniec, S. Baud, B. Dubreucq, J.P. Renou, C. Camilleri, and O. Loudet. 2012. Expression variation in connected recombinant populations of *Arabidopsis thaliana* highlights distinct transcriptome architectures. BMC Genomics 13:117. doi:10.1186/1471-2164-13-117

Dash, S., J. Van Hemert, L. Hong, R.P. Wise, and J.A. Dickerson. 2012. PLEXdb: Gene expression resources for plants and plant pathogens. Nucleic Acids Res. 40:D1194–D1201. doi:10.1093/nar/gkr938

Delourme, R., C. Falentin, V. Huteau, V. Clouet, R. Horvais, B. Gandon, S. Specel, L. Hanneton, J.E. Dheu, M. Deschamps, E. Margale, P. Vincourt, and M. Renard. 2006. Genetic control of oil content in oilseed rape (*Brassica napus* L.). Theor. Appl. Genet. 113:1331–1345. doi:10.1007/ s00122-006-0386-z

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608– 612. doi:10.1007/BF00226905

Drost, D.R., C.I. Benedict, A. Berg, E. Novaes, C.R. Novaes, Q. Yu, C. Dervinis, J.M. Maia, J. Yap, B. Miles, and M. Kirst. 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. doi:10.1073/pnas.0914709107

Druka, A., E. Potokina, Z. Luo, N. Bonar, I. Druka, L. Zhang, D.F. Marshall, B.J. Steffenson, T.J. Close, R.P. Wise, A. Kleinhofs, R.W. Williams, M.J. Kearsey, and R. Waugh. 2008. Exploiting regulatory variation to identify genes underlying quantitative resistance to the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* in barley. Theor. Appl. Genet. 117:261–272. doi:10.1007/s00122-008-0771-x

Fasoula, V.A., D.K. Harris, and H.R. Boerma. 2004. Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. Crop Sci. 44:1218–1225. doi:10.2135/ cropsci2004.1218

Fisher, R.A. 1925. Statistical methods for research workers. Oliver and Boyd, Edinburgh, UK.

Footitt, S., S.P. Slocombe, V. Larner, S. Kurup, Y. Wu, T. Larson, I. Graham, A. Baker, and M. Holdsworth. 2002. Control of germination and lipid mobilization by *COMATOSE*, the Arabidopsis homologue of human ALDP. EMBO J. 21:2912–2922. doi:10.1093/emboj/cdf300

Grupe, A., S. Germer, J. Usuka, D. Aud, J.K. Belknap, R.F. Klein, M.K. Ahluwalia, R. Higuchi, and G. Peltz. 2001. In silico mapping of complex disease-related traits in mice. Science 292:1915–1918. doi:10.1126/science. 1058889

Hammond, J.P., S. Mayes, H.C. Bowen, N.S. Graham, R.M. Hayden, C.G. Love, W.P. Spracklen, J. Wang, S.J. Welham, P.J. White, G.J. King, and M.R. Broadley. 2011. Regulatory hotspots are associated with plant gene expression under varying soil phosphorus supply in *Brassica rapa*. Plant Physiol. 156:1230–1241. doi:10.1104/pp.111.175612

Hobbs, D.H., J.E. Flintham, and M.J. Hills. 2004. Genetic control of storage oil synthesis in seeds of Arabidopsis. Plant Physiol. 136:3341–3349. doi:10.1104/pp.104.049486

Huang, A.H. 1996. Oleosins and oil bodies in seeds and other organs. Plant Physiol. 110:1055–1061. doi:10.1104/pp.110.4.1055

Hyten, D.L., S.B. Cannon, Q. Song, N. Weeks, E.W. Fickus, R.C. Shoemaker, J.E. Specht, A.D. Farmer, G.D. May, and P.B. Cregan. 2010. High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11:38. doi:10.1186/1471-2164-11-38

Hyten, D.L., V.R. Pantalone, C.E. Sams, A.M. Saxton, D. Landau-Ellis, T.R. Stefaniak, and M.E. Schmidt. 2004. Seed quality QTL in a prominent soybean population. Theor. Appl. Genet. 109:552–561. doi:10.1007/ s00122-004-1661-5

Jansen, R.C., and J.P. Nap. 2001. Genetical genomics: The added value from segregation. Trends Genet. 17:388–391. doi:10.1016/S0168- 9525(01)02310-1

Kabelka, E.A., B.W. Diers, W.R. Fehr, A.R. LeRoy, I.C. Baianu, T. You, D.J. Neece, and R.L. Nelson. 2004. Putative alleles for increased yield from soybean plant introductions. Crop Sci. 44:784–791. doi:10.2135/cropsci2004.0784

Keurentjes, J.J., J. Fu, I.R. Terpstra, J.M. Garcia, G. van den Ackerveken, L.B. Snoek, A.J. Peeters, D. Vreugdenhil, M. Koornneef, and R.C. Jansen. 2007. Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc. Natl. Acad. Sci. U.S.A. 104:1708–1713. doi:10.1073/pnas.0610429104

Kirst, M., C.J. Basten, A.A. Myburg, Z.B. Zeng, and R.R. Sederoff. 2005. Genetic architecture of transcript-level variation in differentiating xylem of a eucalyptus hybrid. Genetics 169:2295–2303. doi:10.1534/ genetics.104.039198

Kliebenstein, D.J., M.A. West, H. van Leeuwen, K. Kim, R.W. Doerge, R.W. Michelmore, and D.A. St. Clair. 2006. Genomic survey of gene expression diversity in *Arabidopsis thaliana*. Genetics 172:1179–1189. doi:10.1534/genetics.105.049353

Lark, K.G., K. Chase, F. Adler, L.M. Mansur, and J.H. Orf. 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc. Natl. Acad. Sci. USA 92:4656–4660. doi:10.1073/pnas.92.10.4656

Lark, K.G., J.H. Orf, and L.M. Mansur. 1994. Epistatic expression of quantitative trait loci (QTL) in soybean [*Glycine max* (L.) Merr.] determined by QTL association with RFLP alleles. Theor. Appl. Genet. 88:486–489.

Leach, L.J., Z. Zhang, C. Lu, M.J. Kearsey, and Z. Luo. 2007. The role of *cis*-regulatory motifs and genetical control of expression in the divergence of yeast duplicate genes. Mol. Biol. Evol. 24:2556–2565. doi:10.1093/molbev/msm188

Lee, S.H., M.A. Bailey, M.A.R. Mian, T.E. Carter, Jr., E.R. Shipe, D.A. Ashley, W.A. Parrot, R.S. Hussey, and H.R. Boerma. 1996. RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor. Appl. Genet. 93:649–657. doi:10.1007/BF00224058

Li, L., K. Petsch, R. Shimizu, S. Liu, W.W. Xu, K. Ying, J. Yu, M.J. Scanlon, P.S. Schnable, M.C. Timmermans, N.M. Springer, and G.J. Muehlbauer. 2013. Mendelian and non-Mendelian regulation of gene expression in maize. PLoS Genet. 9:e1003202. doi:10.1371/journal.pgen.1003202

Lin, J.Y., R.M. Stupar, C. Hans, D.L. Hyten, and S.A. Jackson. 2011. Structural and functional divergence of a 1-Mb duplicated region in the soybean (*Glycine max*) genome and comparison to an orthologous region from *Phaseolus vulgaris*. Plant Cell 22:2545–2561.

Liu, P.P., N. Koizuka, R.C. Martin, and H. Nonogaki. 2005. The *BME3* (Blue Micropylar End 3) GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed germination. Plant J. 44:960–971. doi:10.1111/j.1365-313X.2005.02588.x

Liu, S., K. Ying, C.T. Yeh, J. Yang, R. Swanson-Wagner, W. Wu, T. Richmond, D.J. Gerhardt, J. Lai, N. Springer, D. Nettleton, J.A. Jeddeloh, and P.S. Schnable. 2012. Changes in genome content generated via segregation of non-allelic homologs. Plant J. 72:390–399. doi:10.1111/j.1365-313X.2012.05087.x

Mansur, L.M., K.G. Lark, H. Kross, and A. Oliveira. 1993. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). Theor. Appl. Genet. 86:907–913. doi:10.1007/BF00211040

Mehrabian, M., H. Allayee, J. Stockton, P.Y. Lum, T.A. Drake, L.W. Castellani, M. Suh, C. Armour, S. Edwards, J. Lamb, A.J. Lusis, and E.E. Schadt. 2005. Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. Nat. Genet. 37:1224–1233. doi:10.1038/ng1619

Morley, M., C.M. Molony, T.M. Weber, J.L. Devlin, K.G. Ewens, R.S. Spielman, and V.G. Cheung. 2004. Genetic analysis of genome-wide variation in human gene expression. Nature 430:743–747. doi:10.1038/nature02797

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. doi:10.1371/journal.pgen.1002208

Mou, Z., Y. He, Y. Dai, X. Liu, J. Li. 2000. Deficiency in fatty acid synthase leads to premature cell death and dramatic alterations in plant morphology. Plant Cell. 12(3):494–498.

Nyathi, Y., C. De Marcos Lousa, C.W. van Roermund, R.J. Wanders, B. Johnson, S.A. Baldwin, F.L. Theodoulou, and A. Baker. 2010. The Arabidopsis peroxisomal ABC transporter, Comatose, complements the *Saccharomyces cerevisiae pxa1 pxa2*Δ mutant for metabolism of long-chain fatty acids and exhibits fatty acyl-CoA-stimulated ATPase activity. J. Biol. Chem. 285:29892–29902. doi:10.1074/jbc.M110.151225

Orf, J.H., K. Chase, T. Jarvik, L.M. Mansur, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999. Genetics of soybean agronomic traits I. Comparison of three related recombinant inbred populations. Crop Sci. 39:1642–1651. doi:10.2135/cropsci1999.3961642x

Panthee, D.R., V.R. Pantalone, D.R. West, A.M. Saxton, and C.E. Sams. 2005. Quantitative trait loci for seed protein and oil concentration and seed size in soybean. Crop Sci. 45:2015–2022. doi:10.2135/cropsci2004.0720

Parthibane, V., R. Iyappan, A. Vijayakumar, V. Venkateshwari, and R. Rajasekharan. 2012b. Serine/Threonine/Tyrosine protein kinase phosphorylates oleosin, a regulator of lipid metabolic functions. Plant Physiol. 159:95–104. doi:10.1104/pp.112.197194

Parthibane, V., S. Rajakumari, V. Venkateshwari, R. Iyappan, and R. Rajasekharan. 2012a. Oleosin is bifunctional enzyme that has both monoacylglycerol acyltransferase and phospholipase activities. J. Biol. Chem. 287:1946–1954. doi:10.1074/jbc.M111.309955

Potokina, E., A. Druka, Z. Luo, R. Wise, R. Waugh, and M. Kearsey. 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. doi:10.1111/j.1365-313X.2007.03315.x

Quigley, D.A., M.D. To, J. Perez-Losada, F.G. Pelorosso, J.H. Mao, H. Nagase, D.G. Ginzinger, and A. Balmain. 2009. Genetic architecture of mouse skin inflammation and tumour susceptibility. Nature 458:505– 508. doi:10.1038/nature07683

Rockman, M.V., and L. Kruglyak. 2006. Genetics of global gene expression. Nat. Rev. Genet. 7:862–872. doi:10.1038/nrg1964

Ruuska, S.A., T. Girke, C. Benning, and J.B. Ohlrogge. 2002. Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell 14:1191–1206. doi:10.1105/tpc.000877

Ruuska, S.A., J. Schwender, and J.B. Ohlrogge. 2004. The capacity of green oilseeds to utilize photosynthesis to drive biosynthetic processes. Plant Physiol. 136:2700–2709. doi:10.1104/pp.104.047977

Schadt, E.E., S.A. Monks, T.A. Drake, A.J. Lusis, N. Che, V. Colinayo, T.G. Ruff, S.B. Milligan, J.R. Lamb, G. Cavet, P.S. Linsley, M. Mao, R.B. Stoughton, and S.H. Friend. 2003. Genetics of gene expression surveyed in maize, mouse, and man. Nature 422:297–302. doi:10.1038/nature01434

Schmutz, J., S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng, D. Xu, U. Hellsten, G.D. May, Y. Yu, T. Sakurai, T. Umezawa, M.K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X.C. Zhang, K. Shinozaki, H.T. Nguyen, R.A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R.C. Shoemaker, and S.A. Jackson. 2010. Genome sequence of the palaeopolyploid soybean. Nature 463:178–183. doi:10.1038/nature08670

Severin, A.J., G.A. Peiffer, W.W. Xu, D.L. Hyten, B. Bucciarelli, J.A. O'Rourke, Y.T. Bolon, D. Grant, A.D. Farmer, G.D. May, C.P. Vance, R.C. Shoemaker, and R.M. Stupar. 2010a. An integrative approach to genomic introgression mapping. Plant Physiol. 154:3–12. doi:10.1104/pp.110.158949

Severin, A.J., J.L. Woody, Y.T. Bolon, B. Joseph, B.W. Diers, A.D. Farmer, G.J. Muehlbauer, R.T. Nelson, D. Grant, J.E. Specht, M.A. Graham, S.B. Cannon, G.D. May, C.P. Vance, and R.C. Shoemaker. 2010b. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. BMC Plant Biol. 10:160. doi:10.1186/1471-2229-10-160

Shannon, P., A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 13:2498–2504. doi:10.1101/gr.1239303

Shoemaker, R.C., J. Schlueter, and J.J. Doyle. 2006. Paleopolyploidy and gene duplication in soybean and other legumes. Curr. Opin. Plant Biol. 9:104–109. doi:10.1016/j.pbi.2006.01.007

Siloto, R.M., K. Findlay, A. Lopez-Villalobos, E.C. Yeung, C.L. Nykiforuk, and M.M. Moloney. 2006. The accumulation of oleosins determines the size of seed oilbodies in Arabidopsis. Plant Cell 18:1961–1974. doi:10.1105/tpc.106.041269

Sonderby, I.E., B.G. Hansen, N. Bjarnholt, C. Ticconi, B.A. Halkier, and D.J. Kliebenstein. 2007. A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. PLoS ONE 2:e1322. doi:10.1371/journal.pone.0001322

Song, H., Z. Yin, M. Chao, L. Ning, D. Zhang, and D. Yu. 2013. Functional properties and expression quantitative trait loci for phosphate transporter GmPT1 in soybean. Plant Cell Environ. (In press) doi:10.1111/pce.12170

Stombaugh, S.K., J.H. Orf, H.G. Jung, K. Chase, K.G. Lark, and D.A. Somers. 2004. Quantitative trait loci associated with cell wall polysaccharides in soybean seed. Crop Sci. 44:2101–2106. doi:10.2135/cropsci2004.2101

Storey, J.D., and R. Tibshirani. 2003. Statistical significance for genome-wide studies. Proc. Natl. Acad. Sci. U.S.A. 100:9440–9445. doi:10.1073/ pnas.1530509100

Tajuddin, T., S. Watanabe, N. Yamanaka, and K. Harada. 2003. Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines. Breed. Sci. 53:133–140. doi:10.1270/jsbbs.53.133

Thimm, O., O. Blasing, Y. Gibon, A. Nagel, S. Meyer, P. Kruger, J. Selbig, L.A. Muller, S.Y. Rhee, and M. Stitt. 2004. MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 37:914–939. doi:10.1111/j.1365-313X.2004.02016.x

Tuteja, J.H., S.J. Clough, W.C. Chan, and L.O. Vodkin. 2004. Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. Plant Cell 16:819–835. doi:10.1105/tpc.021352

Tuteja, J.H., G. Zabala, K. Varala, M. Hudson, and L.O. Vodkin. 2009. Endogenous, tissue-specific short interfering RNAs silence the chalcone synthase gene family in *Glycine max* seed coats. Plant Cell 21:3063– 3077. doi:10.1105/tpc.109.069856

Van Ooijen, J.W. 1999. LOD significance thresholds for QTL analysis in experimental populations of diploid species. Heredity 83:613–624. doi:10.1038/sj.hdy.6886230

Vuylsteke, M., F. van Eeuwijk, P. Van Hummelen, M. Kuiper, and M. Zabeau. 2005. Genetic analysis of variation in gene expression in *Arabidopsis thaliana*. Genetics 171:1267–1275. doi:10.1534/genetics.105.041509

Wang, C.-S., and L.O. Vodkin. 1994. Extraction of RNA from tissues containing high levels of procyanidins that bind RNA. Plant Mol. Biol. Rep. 12:132–145. doi:10.1007/BF02668374

Wentzell, A.M., I. Boeye, Z. Zhang, and D.J. Kliebenstein. 2008. Genetic networks controlling structural outcome of glucosinolate activation across development. PLoS Genet. 4:e1000234. doi:10.1371/journal.pgen.1000234

Wentzell, A.M., H.C. Rowe, B.G. Hansen, C. Ticconi, B.A. Halkier, and D.J. Kliebenstein. 2007. Linking metabolic QTLs with network and *cis*-eQTLs controlling biosynthetic pathways. PLoS Genet. 3:1687–1701. doi:10.1371/journal.pgen.0030162

West, M.A., K. Kim, D.J. Kliebenstein, H. van Leeuwen, R.W. Michelmore, R.W. Doerge, and D.A. St. Clair. 2007. Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics 175:1441–1450. doi:10.1534/genetics.106.064972

Wu, C., D.L. Delano, N. Mitro, S.V. Su, J. Janes, P. McClurg, S. Batalov, G.L. Welch, J. Zhang, A.P. Orth, J.R. Walker, R.J. Glynne, M.P. Cooke, J.S. Takahashi, K. Shimomura, A. Kohsaka, J. Bass, E. Saez, T. Wiltshire, and A.I. Su. 2008. Gene set enrichment in eQTL data identifies novel annotations and pathway regulators. PLoS Genet. 4:e1000070. doi:10.1371/journal.pgen.1000070

Yin, Z., F. Meng, H. Song, X. Wang, M. Chao, G. Zhang, X. Xu, D. Deng, and D. Yu. 2011. GmFtsH9 expression correlates with in vivo Photosystem II function: Chlorophyll a fluorescence transient analysis and eQTL mapping in soybean. Planta 234:815–827. doi:10.1007/s00425-011-1445-5

Yin, Z., F. Meng, H. Song, X. Wang, X. Xu, and D. Yu. 2010. Expression quantitative trait loci analysis of two genes encoding rubisco activase in soybean. Plant Physiol. 152:1625–1637. doi:10.1104/pp.109.148312

Zeng, Z.B. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457–1468.

Zhao, J., J. Huang, F. Chen, F. Xu, X. Ni, H. Xu, Y. Wang, C. Jiang, H. Wang, A. Xu, R. Huang, D. Li, and J. Meng. 2012. Molecular mapping of *Arabidopsis thaliana* lipid-related orthologous genes in *Brassica napus*. Theor. Appl. Genet. 124:407–421. doi:10.1007/s00122-011-1716-3

Zheng, Q., and X.J. Wang. 2008. GOEAST: A web-based software toolkit for gene ontology enrichment analysis. Nucleic Acids Res. 36:W358–W363. doi:10.1093/nar/gkn276

TPG13-08-0027 Supplemental Table/figure captions:

SUPPLEMENTAL MATERIAL

Bolon_SuppFigs.pdf

Supplemental Figure S1: Soybean seeds and pod at early seed maturation stage. Immature green soybean seed tissue was collected from each genotype during early seed maturation for the transcript accumulation dataset. A) Image of the immature seed within the pod shell. B) The immature seed harvested with the seed coat. C) Image of seeds attached to the inner pod shell. D) View of the immature green seed with partial seed coat removed.

Supplemental Figure S2: Enrichment of photosynthesis and fatty acid biosynthesis and metabolism Gene Ontology categories among genes with eQTL at the chromosome 20 eQTL hotspot is shown using GOEAST (Zheng and Wang, 2008).

Supplemental Figure S3: Chromosome 20 (~2498 total cM) eQTL hotspot. Genes with eQTL at the ~2498 cM hotspot are represented by colored squares in different metabolic categories using Mapman (Thimm et al., 2004). The color of each square indicates the allele (red = Minsoy, blue = Noir 1) with additive effect, and the scale is based on LOD score for the eQTL. Twelve eQTL are for genes annotated to lipid biosynthesis with the same additive allele ('Minsoy').

Supplemental Figure S4: Network of eQTL on chromosome 7 (797 cM) and chromosome 20 (2493 cM) mapped for a common set of genes. Edges connecting the nodes are colored red for upregulation of gene transcript accumulation by the 'Minsoy' allele and blue for upregulation of transcription factor gene transcript accumulation by the 'Noir 1' allele. Edge thickness represents LOD (thicker edge = higher LOD). A zigzag edge represents a *cis*-acting regulatory connection

vs. a *trans*-acting regulatory connection. Supplemental Table S12 contains the key for the genes with eQTL depicted in this figure.

Supplemental Figure S5: Enrichment of photosynthesis and metabolism gene ontology categories among genes with eQTL at the chromosome 7 eQTL hotspot is shown using GOEAST (Zheng and Wang, 2008).

Supplemental Figure S6: Phenotypic QTL mapped in MxN RILs. (A-B) Seed oil and protein QTL traces. C) MxN RILs graph of genome-wide eQTL frequency.

Supplemental Table S1: MxN RILs genotype data for 557 Goldengate SNP markers.

Bolon_TableS1.xls

Supplemental Table S2: Genetic map (~2500 cM) of MxN with 557 SNP markers across 20 chromosomes.

Bolon_TableS2.xls

Supplemental Table S3: Seed Oil and Protein QTL mapped in Minsoy x Noir 1. Three QTL with LOD>3.5 were mapped for seed protein in multiple years for the Minsoy x Noir 1 RIL population that included one QTL on chromosome 8 and two QTL for seed protein on chromosomes 4 and 6. Two of these seed protein QTL, on chromosomes 6 and 8, correlated inversely with seed oil QTL.

Bolon_TableS3.xls

Supplemental Table S4: List of 200 gene expression probeset values and annotations for genes with differentially accumulated transcript levels between 'Minsoy' and 'Noir 1' in stage zero of the developing soybean seed.

Bolon_TableS4.xls

Supplemental Table S5: Final normalized gene expression data from Soybean Affymetrix GeneChips on 93 MxN RILs.

Bolon_TableS5.xls

Supplemental Table S6: Genes with eQTL within the chromosome 8 eQTL hotspot enriched for flavonoid biosynthesis pathway genes.

Bolon_TableS6.xls

Supplemental Table S7: eQTL data in the MxN RILs for seed-specific genes shown in Figure 3A.

Bolon_TableS7.xls

Supplemental Table S8: eQTL data for photosynthesis genes, fatty acid biosynthesis genes and oleosin genes shown in Figures 3B-3D.

Bolon_TableS8.xls

Supplemental Table S9: Genes with eQTL peaks at the 2498 total cM eQTL hotspot on chromosome 20 enriched for seed-specific genes, photosynthesis genes, fatty acid biosynthesis-related genes and oleosin genes.

Bolon_TableS9.xls

Supplemental Table S10: Genes with eQTL peaks at the 797 total cM eQTL hotspot on chromosome 7 enriched for photosynthesis genes and metabolic pathway genes.

Bolon_TableS10.xls

Supplemental Table S11: Photosynthesis, fatty acid biosynthesis, oleosin, and other seed-specific genes with eQTL depicted in the network among the three eQTL hotspots on chromosome 7, 13, and 20 in Figure 4.

Bolon_TableS11.xls

Supplemental Table S12: Genes with eQTL depicted in the network between chromosome 7 (797 cM) and chromosome 20 (2493 cM) in Supplemental Figure S4.

Bolon_TableS12.xls

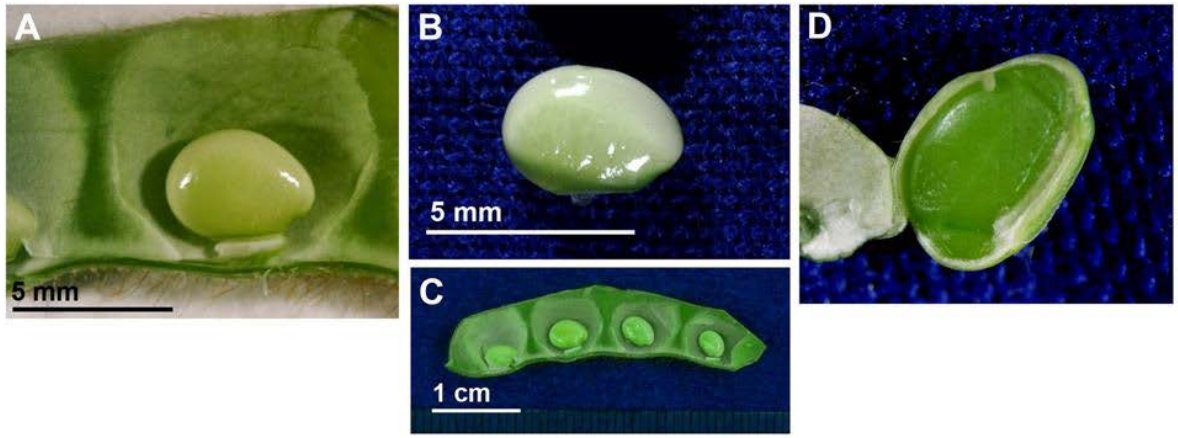
Supplemental Table S13: RNA-seq data in LoPro and HiPro for genes within the chromosome 20 hotspot, including eight seed-specific genes and 36 transcription factor genes in Figure 6.

Bolon_TableS13.xls

Supplemental Table S14: Genes with *cis*-acting eQTL peaks that coincide with major seed protein and oil QTL peaks.

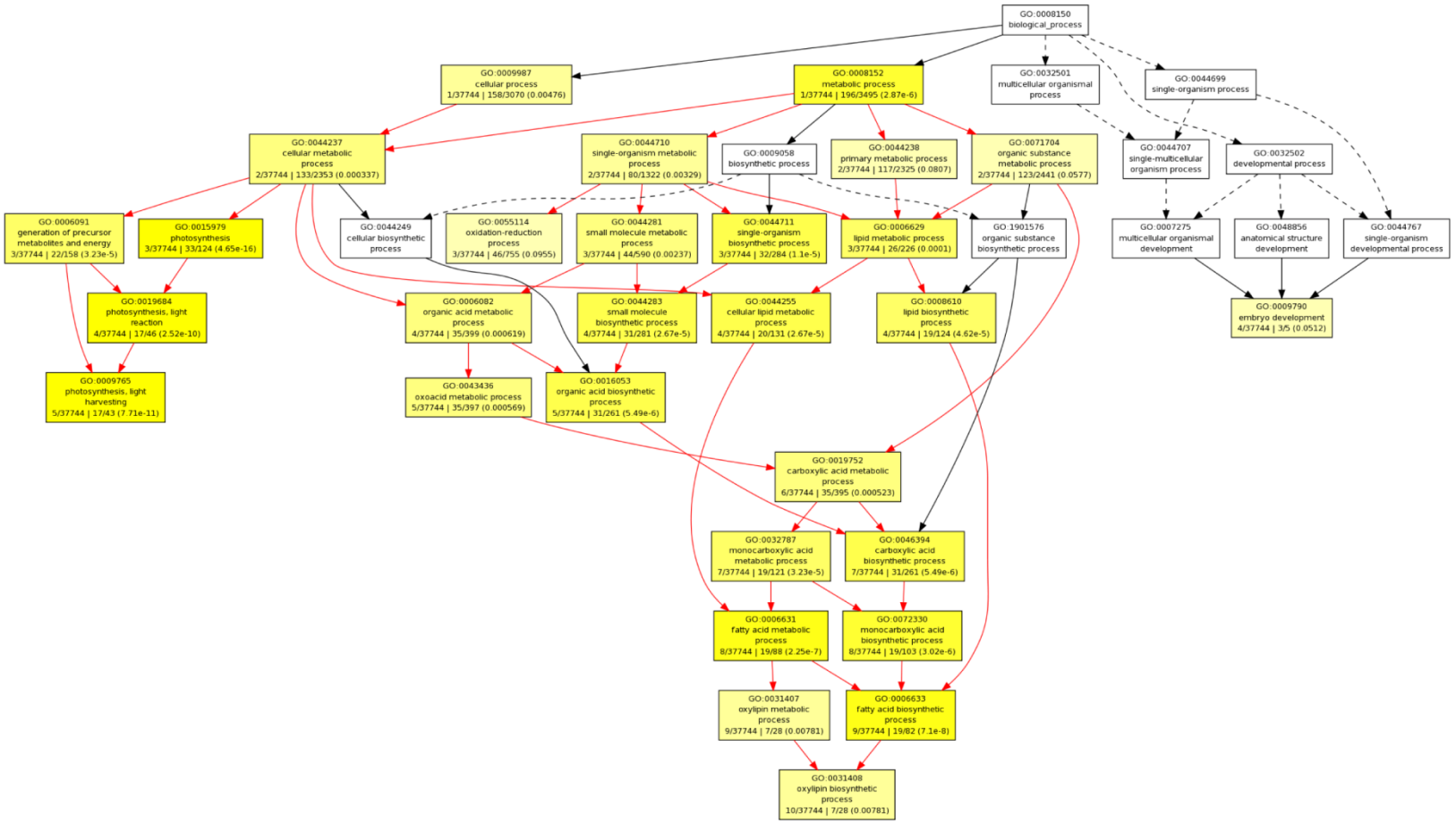
Bolon_TableS14.xls

The Plant Genome: doi: 10.3835/plantgenome2013.08.0027



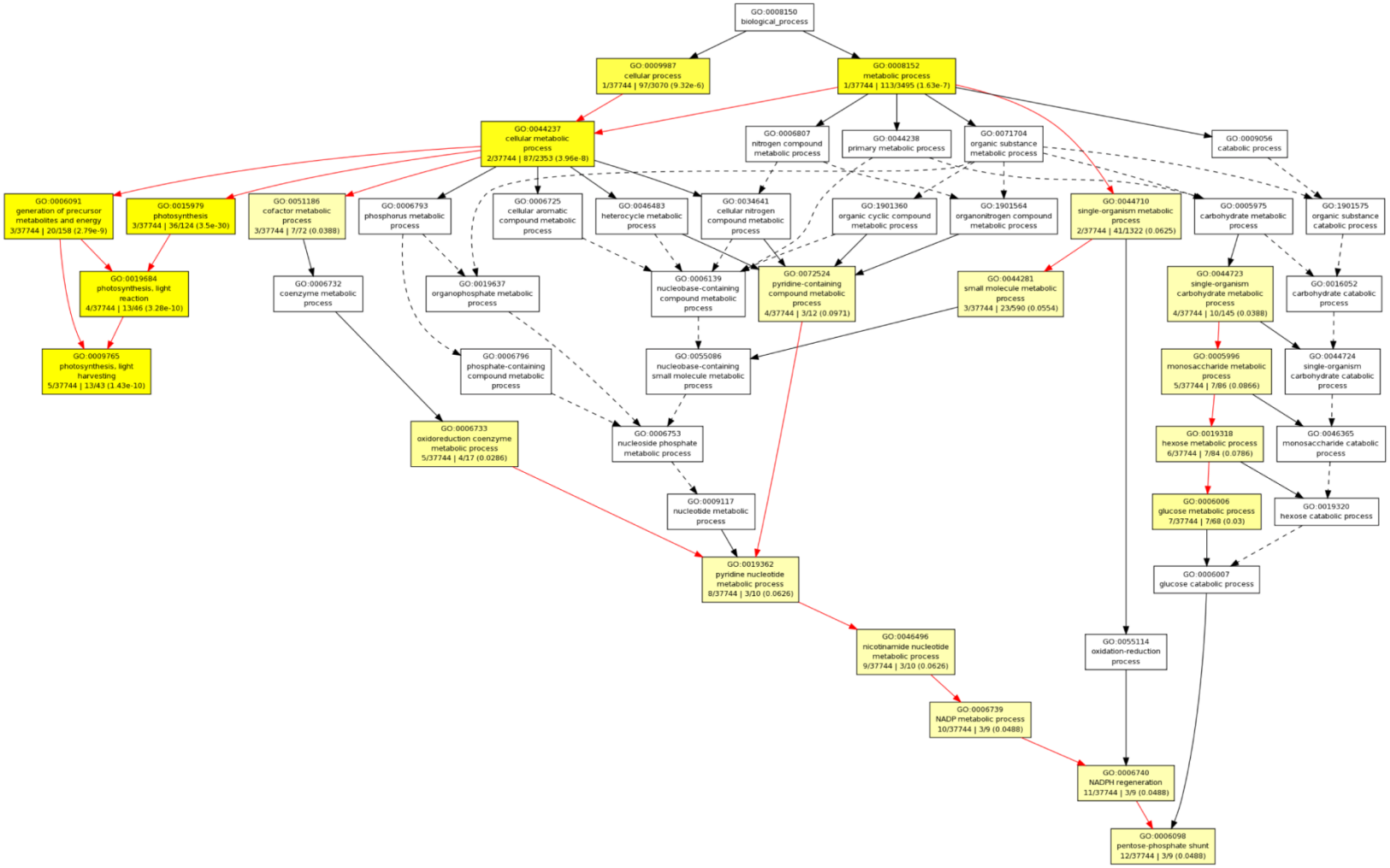
Supplemental Figure S1. Soybean seeds and pod at early seed maturation stage. Immature green soybean seed tissue was collected from each genotype during early seed maturation for the transcript accumulation dataset. (A) Image of the immature seed within the pod shell. (B) The immature seed harvested with the seed coat. (C) Image of seeds attached to the inner pod shell. (D) View of the immature green seed with partial seed coat removed.

Figure S2



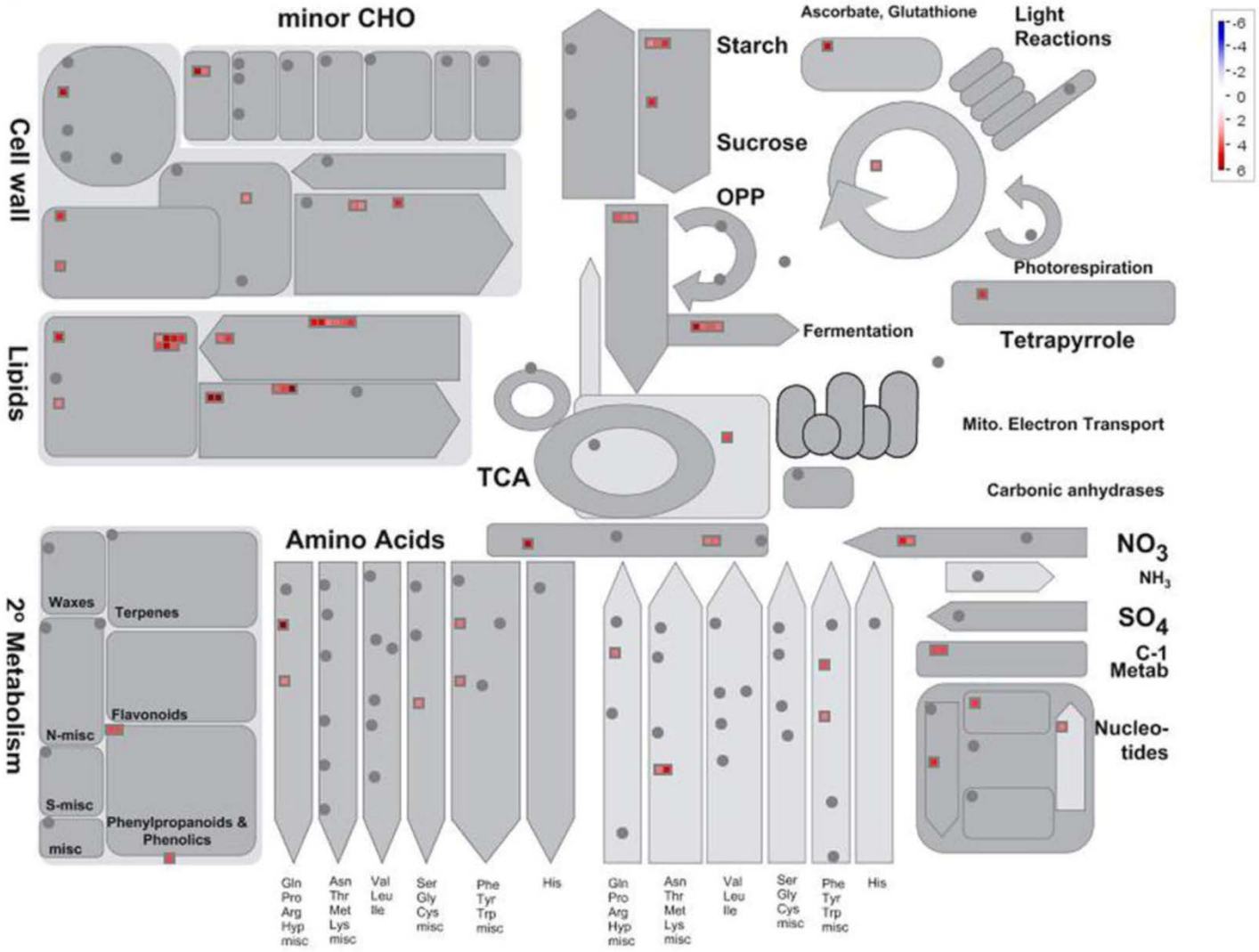
Supplemental Figure S2. Enrichment of photosynthesis and fatty acid biosynthesis and metabolism Gene Ontology categories among genes with eQTL at the chromosome 20 eQTL hotspot is shown using GOEAST (Zheng and Wang, 2008).

Figure S3



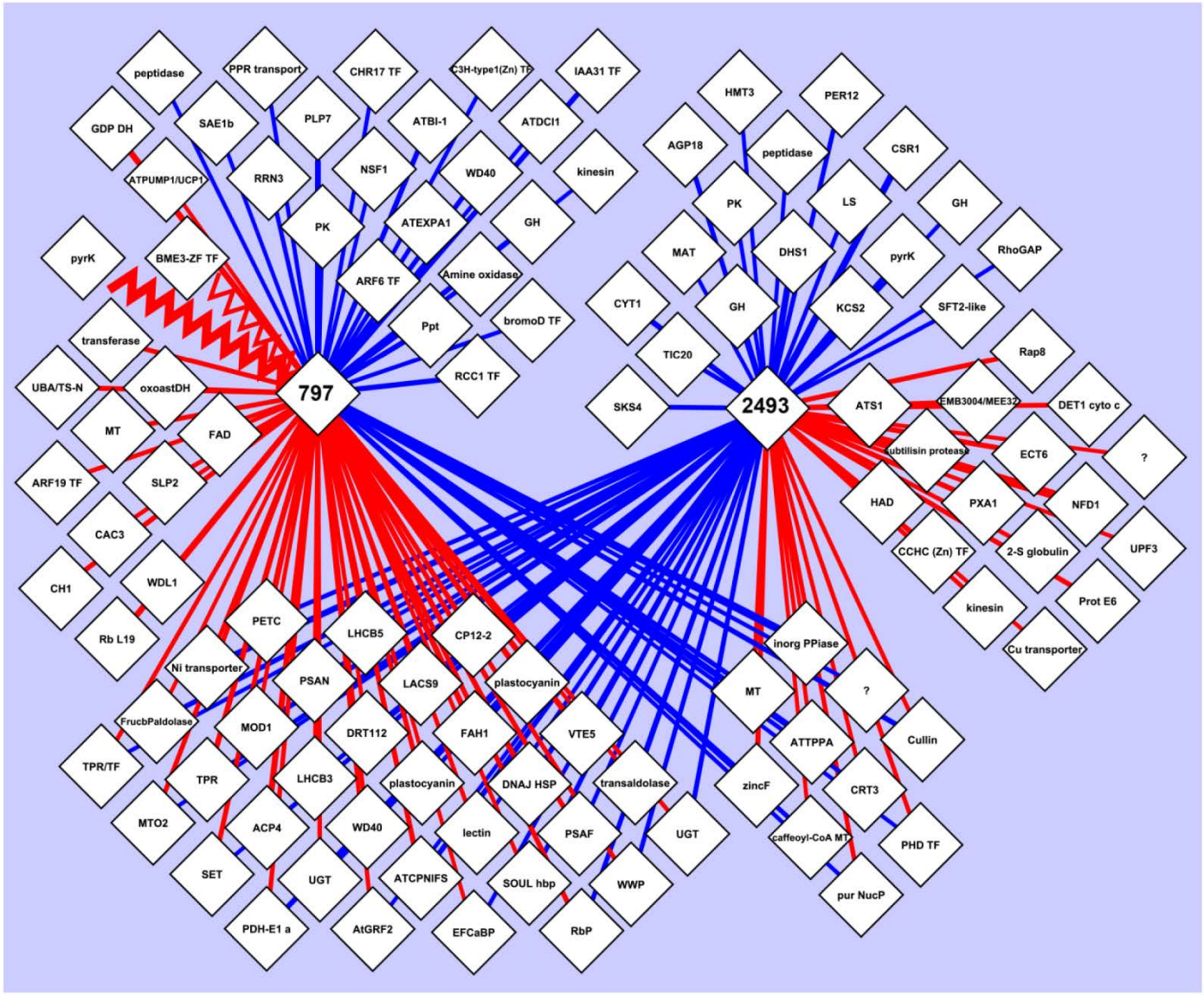
Supplemental Figure S3. Enrichment of photosynthesis and metabolism gene ontology categories among genes with eQTL at the chromosome 7 eQTL hotspot is shown using GOEAST (Zheng and Wang, 2008).

Figure S4


minor CHO
Cell wall
Lipids
2° Metabolism
Waxes
Terpenes
Flavonoids
N-misc
S-misc
misc
Phenylpropanoids & Phenolics
Starch
Sucrose
OPP
Fermentation
TCA
Ascorbate, Glutathione
Light Reactions
Photorespiration
Tetrapyrrole
Mito. Electron Transport
Carbonic anhydrases
Amino Acids
Gln Pro Arg Hyp misc
Asn Thr Met Lys misc
Val Leu Ile
Ser Gly Cys misc
Phe Tyr Trp misc
His
NO3
NH3
SO4
C-1 Metab
Nucleo-tides
-6
-4
-2
0
2
4
6

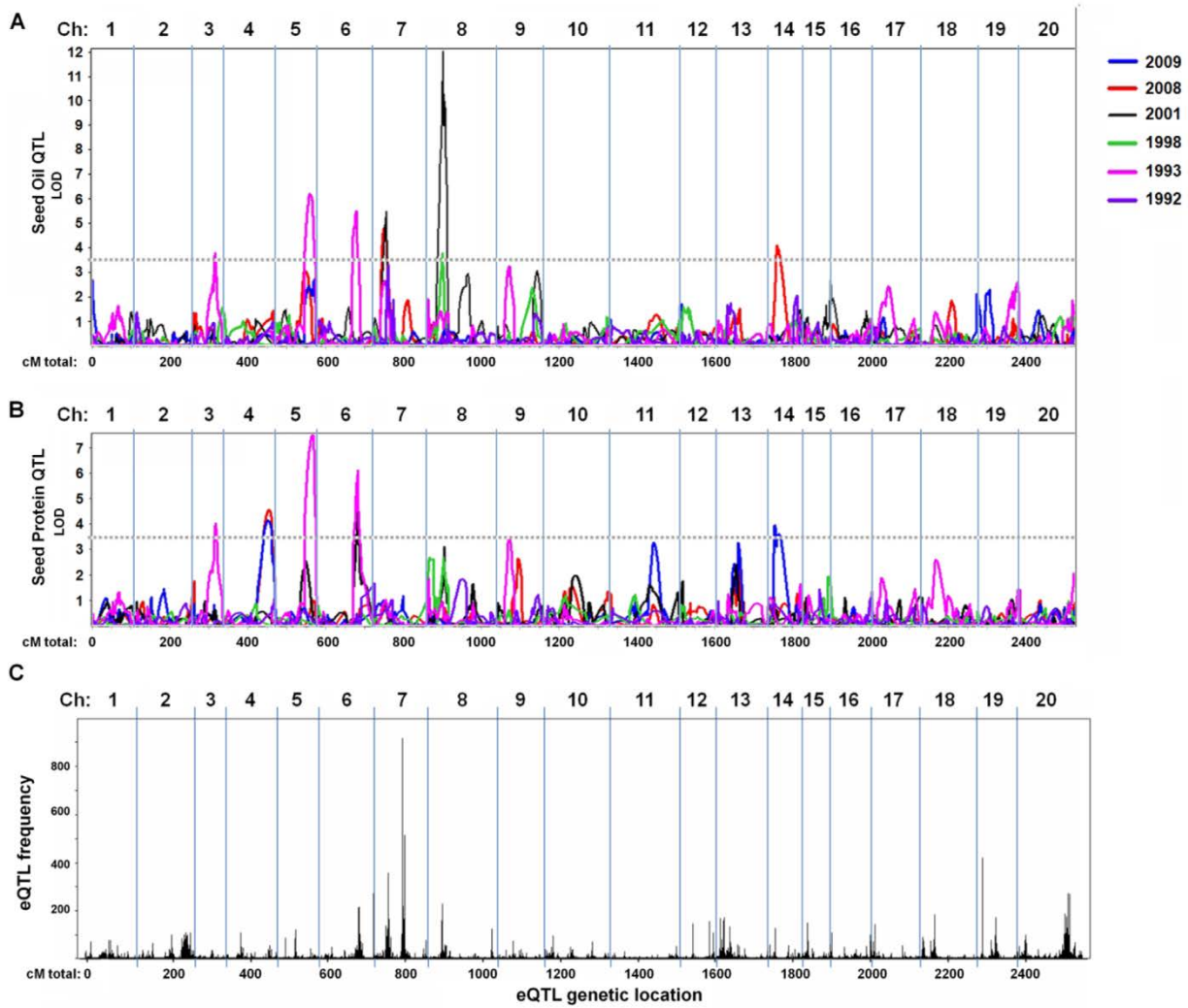
Supplemental Figure S4. Chromosome 20 (~2498 total cM) eQTL hotspot. Genes with eQTL at the ~2498 cM hotspot are represented by colored squares in different metabolic categories using Mapman (Thimm et al., 2004). The color of each square indicates the allele (red = Minsoy, blue = Noir 1) with additive effect, and the scale is based on LOD score for the eQTL. Twelve eQTL are for genes annotated to lipid biosynthesis with the same additive allele ('Minsoy').

Figure S5


797
2493
peptidase
PPR transport
CHR17 TF
C3H-type1(Zn) TF
IAA31 TF
GDP DH
SAE1b
PLP7
ATBI-1
ATDCI1
ATPUMP1/UCP1
RRN3
NSF1
WD40
kinesin
PK
ATEXPA1
GH
pyrK
BME3-ZF TF
ARF6 TF
Amine oxidase
transferase
Ppt
bromoD TF
UBA/TS-N
oxoastDH
RCC1 TF
MT
FAD
ARF19 TF
SLP2
CAC3
CH1
WDL1
Rb L19
PETC
LHCB5
CP12-2
Ni transporter
PSAN
LACS9
plastocyanin
FrucbPaldolase
MOD1
DRT112
FAH1
VTE5
TPR/TF
TPR
LHCB3
plastocyanin
DNAJ HSP
transaldolase
MTO2
ACP4
WD40
lectin
PSAF
UGT
SET
UGT
ATCPNIFS
SOUL hbp
WWP
PDH-E1 a
AtGRF2
EFCaBP
RbP
HMT3
PER12
AGP18
peptidase
CSR1
PK
LS
GH
MAT
DHS1
pyrK
RhoGAP
CYT1
GH
KCS2
SFT2-like
TIC20
SKS4
Rap8
ATS1
EMB3004/MEE32
DET1 cyto c
subtilisin protease
ECT6
?
HAD
PXA1
NFD1
CCHC (Zn) TF
2-S globulin
UPF3
kinesin
Prot E6
Cu transporter
inorg PPiase
MT
?
ATTPPA
Cullin
zincF
CRT3
caffeoyl-CoA MT
PHD TF
pur NucP

Supplemental Figure S5. Network of eQTL on chromosome 7 (797 cM) and chromosome 20 (2493 cM) mapped for a common set of genes. Edges connecting the nodes are colored red for upregulation of gene transcript accumulation by the '*Minsoy'* allele and blue for upregulation of transcription factor gene transcript accumulation by the 'Noir 1' allele. Edge thickness represents LOD (thicker edge = higher LOD). A zigzag edge represents a *cis*-acting regulatory connection vs. a *trans*-acting regulatory connection. Supplemental Table S12 contains the key for the genes with eQTL depicted in this figure.

Figure S6


A
Ch: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Seed Oil QTL LOD
cM total: 0 200 400 600 800 1000 1200 1400 1600 1800 2000 2200 2400
2009
2008
2001
1998
1993
1992
B
Ch: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Seed Protein QTL LOD
cM total: 0 200 400 600 800 1000 1200 1400 1600 1800 2000 2200 2400
C
Ch: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
eQTL frequency
cM total: 0 200 400 600 800 1000 1200 1400 1600 1800 2000 2200 2400
eQTL genetic location

Supplemental Figure S6. Phenotypic QTL mapped in MxN RILs. (A-B) Seed oil and protein QTL traces. C) MxN RILs graph of genome-wide eQTL frequency.