Identification of Candidate Genes Underlying an Iron Efficiency Quantitative Trait Locus in Soybean

Gregory A. Peiffer, Keith E. King, Andrew J. Severin, Gregory D. May, Silvia R. Cianzio, Shun Fu Lin, Nicholas C. Lauter, and Randy C. Shoemaker*

Department of Agronomy, Iowa State University, Ames, Iowa 50010 (G.A.P., K.E.K., A.J.S., S.R.C.); National Center for Genome Research, Santa Fe, New Mexico 87505 (G.D.M.); Department of Agronomy, National Taiwan University, Taipei, Taiwan, Republic of China (S.F.L.); and Corn Insects and Crop Genetics Research Unit, United States Department of Agriculture-Agricultural Research Service, Ames, Iowa 50010 (N.C.L., R.C.S.)

Prevalent on calcareous soils in the United States and abroad, iron deficiency is among the most common and severe nutritional stresses in plants. In soybean (*Glycine max*) commercial plantings, the identification and use of iron-efficient genotypes has proven to be the best form of managing this soil-related plant stress. Previous studies conducted in soybean identified a significant iron efficiency quantitative trait locus (QTL) explaining more than 70% of the phenotypic variation for the trait. In this research, we identified candidate genes underlying this QTL through molecular breeding, mapping, and transcriptome sequencing. Introgression mapping was performed using two related near-isogenic lines in which a region located on soybean chromosome 3 required for iron efficiency was identified. The region corresponds to the previously reported iron efficiency QTL. The location was further confirmed through QTL mapping conducted in this study. Transcriptome sequencing and quantitative real-time-polymerase chain reaction identified two genes encoding transcription factors within the region that were significantly induced in soybean roots under iron stress. The two induced transcription factors were identified as homologs of the subgroup lb basic helix-loop-helix (bHLH) genes that are known to regulate the strategy I response in Arabidopsis (*Arabidopsis thaliana*). Resequencing of these differentially expressed genes unveiled a significant deletion within a predicted dimerization domain. We hypothesize that this deletion disrupts the Fe-DEFICIENCY-INDUCED TRANSCRIPTION FACTOR (FIT)/bHLH heterodimer that has been shown to induce known iron acquisition genes.

Iron deficiency chlorosis (IDC) is a worldwide concern. In soybean (*Glycine max*), a 20% yield reduction has been reported for each one-point increase in iron chlorosis score (Froehlich and Fehr, 1981). Hansen et al. (2003) estimated an annual loss of $120 million due to IDC in the United States alone. Soybean bushels are selling approximately $7 higher today than in 2004, increasing the estimated current revenue losses to over $260 million.

Iron is one of seven essential micronutrients required for plant growth (Crosa, 1989). If concentrations of micronutrients within the plant are too high, they become toxic and will stunt growth in the same fashion as an insufficiency (Havlín et al., 1999). Iron deficiency is one of the most common micronutrient deficiencies in the world due to alkaline soil conditions (pH > 8) that maintain ferric (Fe³⁺) iron in an insoluble state unavailable to some plant genotypes. Conversely, most plants are able to take up and utilize soluble ferrous (Fe²⁺) iron, prevalent in soils with pH ranging from 6.5 to 7.5 (Havlín et al., 1999). Iron deficiency in plants is commonly an issue of iron availability and not one of iron supply, as iron is the fourth most abundant element on the earth’s surface.

IDC is caused by a decrease in photosynthetic pigments (chlorophylls *a/b*, carotene, xanthophylls, etc.), in addition to a reduced electron transport potential and a significant reduction in the thylakoid system (Spiller and Terry, 1980; Terry, 1980; Taylor et al., 1982). Symptoms of IDC vary from interveinal yellowing to complete necrosis (Cianzio et al., 1979; Havlin et al., 1999), which most genotypes overcome at the end of the growing season; however, the severity of chlorosis is tightly correlated with end-of-season yield losses (Inskeep and Bloom, 1987).

The best way to manage IDC is through the use of iron-efficient varieties in commercial plantings. This is the reason why so much research has been devoted to identifying genetic markers associated with iron efficiency for use in breeding programs. Work by Lin et al. (1997) showed that there are two genetic models to explain IDC resistance at the molecular level, which had been previously identified by classical genetic studies (Cianzio et al., 1980; Cianzio and Fehr, 1982). One model involves a single major gene with modifiers (Cianzio et al., 1980) identified in the Anoka × A7 population; the associated quantitative trait locus (QTL) was located on chromosome 3, explaining...
more than 70% of the phenotypic variation (Lin et al., 1997). The second model involved multiple genes (Cianzio and Fehr, 1982) identified in the PrideB216 × A15 population, which represented a classic quantitative mode of inheritance. Lin et al. (1997) were unable to identify QTLs in this population.

All plants except the Poaceae family have adapted what is known as the strategy I response (Marschner et al., 1986) in order to reduce and acquire Fe$^{3+}$ from soils. In strategy I, plants pump protons across the plasma membrane via an H$^+$/ATPase to acidify the surrounding soil (Fox and Guerinot, 1998). The resultant acidification helps solubilize unavailable ferric iron. A ferric-chelate reductase (FRO2) reduces chelated Fe$^{3+}$ into available Fe$^{2+}$ (Walker and Connolly, 2008). The reduced iron (Fe$^{2+}$) is transported across the plasma membrane via an iron-regulated transporter (IRT1; Korshunova et al., 1999). Recently, work in Arabidopsis (Arabidopsis thaliana) and tomato (Solanum lycopersicum) has made great strides at elucidating plant responses to iron-limiting conditions. The response is initiated by an unknown factor that results from the recognition of low cellular iron, which then activates the transcription factors Arabidopsis thaliana Fe-DEFICIENCY-INDUCED TRANSCRIPTION FACTOR (AtFIT), Arabidopsis thaliana basic helix-loop-helix 038 (AtbHLH038), and AtbHLH039 (Bauer et al., 2007; Wang et al., 2007; Lingam et al., 2011). Yuan et al. (2008) showed that in Arabidopsis, AtFIT/AtbHLH038 or AtFIT/AtbHLH039 was capable of initiating β-glucuronidase (GUS) expression when regulated by AtFRO2 and AtIRT1 promoters. Their results suggested that AtFIT binds as a heterodimer to either AtbHLH038 or AtbHLH039 to initiate the expression of AtFRO2 and AtIRT1 under iron-limiting conditions (Yuan et al., 2008).

Soybean has a production disadvantage when it is unable to evoke an effective iron deficiency stress response. Near-isogenic lines (NILs) differing in their response to iron were developed by the United States Department of Agriculture-Agricultural Research Service in the early 1970s (Bernard, 1975) and have become some of the standard lines for IDC research in soybean. From this work, two lines identified as PI 547430 and PI 547698 resulted from crosses Clark (PI 548533) × T203 (PI 54619) and Harosoy (PI 548573) × T203 (PI 54619), respectively.

Results

Identification of the T203 Introgressions in Two NILs

U.S. Department of Agriculture-Agricultural Research Service scientists generated NILs by crossing the iron-efficient recurrent parents, Clark (PI 548533) and Harosoy (PI 548573), with an inefficient donor parent, T203 (PI 54619). The resulting F2 progeny were screened for iron efficiency and then backcrossed to the recurrent parent for six generations to create the iron-inefficient NIL. Therefore, the genetic material from T203 introgressed into each efficient background presumably contains the gene(s) required to induce an iron-inefficient response in an otherwise efficient background (Bernard, 1975).

Identification of the introgressed DNA within the NIL genome was accomplished through a genome-wide survey using greater than 860 published simple sequence repeat (SSR) markers, available at Soybase.org (composite map 2003). At least one marker representing the allele from the donor parent (T203) was detected on 12 of the 20 chromosomes for each background, Clark and Harosoy. Only on chromosome 3 (Gm03) was an introgressed region common for the two NILs (Table I). PI 547430 and T203 shared alleles for markers Satt387 → Satt125, and PI 547698 and T203 shared alleles for markers Satt275 → Satt339
(Fig. 1). In each NIL, the region contained approximately 37% of all the introgressed markers.

A comparison of the overlapping region between the two pairs of NILs allowed for the region of interest to be narrowed to the 4.2-Mb overlap (Sat387 → Satt339; Fig. 1). Since there were no other overlapping regions on any of the other chromosomes, it is likely that the gene(s) controlling iron efficiency within the Clark/Harosoy NILs resides within this region. The region of introgression overlap on Gm03 correlates with the location of a known iron efficiency QTL in an A7 × Anoka population (Lin et al., 1997, 2000). The iron-inefficient parent of this population, Anoka, is the result of the three-way cross (Lincoln2 × Richland) × Korean. It is important to indicate that Clark, one of the parents of the NILs, also originated from the same single cross Lincoln2 × Richland, later used as a parent in the three-way cross from which Anoka was developed. Although Anoka was not generated in the same fashion as the two NILs, markers screened against Lincoln, Richland, Korean, and Anoka identified a region corresponding to the NIL introgression in which Anoka and Korean shared marker alleles (data not shown), suggesting the transfer of genetic material within this region from the inefficient parent, Korean, to Anoka.

### High-Density Mapping of the Anoka × A7 QTL

Integrating the genetic map developed by Lin et al. (1997) with previously mapped SSRs and newly available BARCSOYSSRs (Song et al., 2010) resulted in a higher density genetic map. The 1993 and 1994 combined visual phenotypic data collected by Lin et al. (1997) were reanalyzed using the integrated map. The QTL on Gm03 was reconfirmed with a logarithm of the odds (LOD) score of 15.29, which is significantly higher than the permutated genome-wise LOD cutoff value of 3.8 (Fig. 2). The QTL region identified in 1997 spanned Gm03 from Sat_33 to BLT_15. By placing additional markers on the map, the predicted QTL region on Gm03 was significantly narrowed using the integrated map (Sat_33 → BARCSOYSSR_03_1129; Fig. 2).

### Reducing the Donor Parent Introgression in PI 547430

Following the identification of the T203 introgressed region in the iron-inefficient NIL PI 547430, further backcrossing to the recurrent parent (Clark) resulted in a BC2 population of PI 547430 during the summer of 2009 (Ames, IA). Two hundred cross-pollinations were performed, and 156 resulted in seed set. Seeds from the crosses were planted in the greenhouse during the winter of 2009/2010. Four SSR markers within the introgressed region were used for screening to distinguish between plants that resulted from crosses or self-pollinated seeds. From these checks, 98 F1 plant/F2 seed families were confirmed as true hybrids. In the summer of 2010, 40 seeds of each confirmed plant/family were planted at Bruner Farm (Ames, IA).

Leaf tissue was harvested early in the growing season, at plant stage V2, in order to genotype the nearly 6,000 plants in the field. An initial round of genotyping was performed with four markers to identify plants with crossovers within theintrogressed region (sub-NILs). From this initial screen, 480 plants were identified with a recombination within the region and chosen for additional marker analysis to identify potentially narrowed introgression regions. Twenty-nine plants were identified as fitting into 10 different subNIL classes, segmenting the introgressed region (Fig. 3).

### Phenotypic Analysis of Lines with Narrowed Introgressions

Phenotyping of the lines representing various recombinant classes was performed during winter 2010/2011. Plants were scored for iron chlorosis symptoms using the standard visual 1 to 5 scale and a Minolta SPAD-502 chlorophyll meter. Two replications, with eight plants each, were grown hydroponically in iron-limiting conditions in the greenhouse. The recombinant classes segregated into two groups, those with introgressions originating from the left side and those originating from the right side of the original PI 547430 introgression (Fig. 3). Visual and SPAD score averages for each of the two groups and controls (Clark and PI 547430) were compared. Interestingly, the visual score and SPAD averages for all categories containing the donor parent allele for Satt387 were not significantly different from the averages for the PI 547430 (P = 0.7366/0.8202) allele. Visual and SPAD averages for all classes containing the Clark allele for Satt387 were not significantly different from the averages of Clark (P =

---

**Table 1. Polymorphic SSR markers screened in the two NILs**

Comparison of polymorphic markers between the two recurrent parents and T203. In each NIL, approximately 37% of the polymorphic markers were clustered on chromosome 3.

<table>
<thead>
<tr>
<th>Parent</th>
<th>Total Markers</th>
<th>Polymorphic Markers</th>
<th>Introgressed Markers</th>
<th>Percentage on Gm03</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clark</td>
<td>869</td>
<td>415</td>
<td>27</td>
<td>37.7</td>
</tr>
<tr>
<td>Harosoy</td>
<td>864</td>
<td>345</td>
<td>19</td>
<td>36.9</td>
</tr>
</tbody>
</table>

*Isoline with the same banding as Clark. **Isoline with the same banding as T203.

---

**Figure 1. Overlay of the two NIL introgressions and the iron efficiency QTL on chromosome 3. The solid box is the previously identified iron efficiency QTL, the wide-dashed box is the PI 547698 introgression, and the narrow-dashed box is the introgression in PI 547430.**
Peiffer et al.

0.0864/0.3051). Of the classes containing the donor parent allele for Satt387, class 1 contains the least amount of T203-derived DNA. Recombinant class 1 was not phenotypically different from the other introgressions containing the T203 allele for Satt387 \((P = 0.097)\). Thus, we could narrow the donor parent introgression in PI 547430 from approximately 6 million bp down to just over 1 million bp.

Fine-Mapping Recombinant Class 1

Recombinant class 1, consisting of the F2.3 family 4-25, contained the least amount of T203 alleles conferring iron inefficiency. Eight plants from the 4-25 family were genotyped with 17 additional SSR markers to identify more precise recombination intervals. Five shared the recombination interval BARCSOYSSR_03_1141\(\rightarrow\)BARCSOYSSR_03_1155 (Fig. 4). Three shared an even narrower recombination interval between BARCSOYSSR_03_1110 and BARCSOYSSR_03_1112 (Fig. 4). The average phenotypic scores for the two subgroups of plants within class 1 were not significantly different from each other \((P = 0.2552)\). Therefore, the introgression was further narrowed to the approximately 250-kb region identified by the three 4-25 plants with the recombination interval BARCSOYSSR_03_1110 \(\rightarrow\)BARCSOYSSR_03_1112. The narrowing of this introgression significantly reduced the number of candidate genes from hundreds of genes down to 18 genes potentially controlling the iron stress response in the NILs.

Next-Generation Sequencing, RNA-Seq, and Analysis of NILs

RNA-Seq was performed in replicate on RNA isolated from Clark leaves and roots 24 h post iron stress. Eight libraries were constructed and sequenced, producing 165,968,390 reads. The reads were then mapped to the soybean gene calls using Bowtie/Tophat (Trapnell et al., 2009). Differential gene expression between iron-sufficient and iron-insufficient conditions was determined using a false discovery rate threshold of 0.05 in Cuffdiff (Roberts et al., 2011). The 18 genes within the narrowed Gm03 introgression were examined in the RNA-Seq data to determine their expression differences between iron-sufficient and iron-insufficient conditions (Table II). In both replicates, only two of the 18 genes were significantly differentially expressed in the roots, Glyma03g28610 and Glyma03g28630. These two genes share 87.8% peptide similarity and are homologous to Arabidopsis AtbHLH038, with 66.7% and 63.9% peptide similarity, and to AtbHLH039, with 55.1% and 58.1% peptide similarity, respectively. The expression of the genes was significantly increased \((P < 0.05)\) under iron stress conditions compared with the iron-sufficient condition. Expression of the two genes was also restricted to the roots, and expression was not detected in leaves of plants grown in stressed or nonstressed conditions. The leaves showed differential expression of six genes within the identified introgression (Glyma03g28450,
Glyma03g28490, Glyma03g28590, Glyma03g28480, Glyma03g28510, and Glyma03g28570) in one of the two biological replicates but not in the other. When the two replicates were combined, four of the genes were confirmed as differentially expressed (Glyma03g28450, Glyma03g28490, Glyma03g28480, and Glyma03g28570; Table II). Within the narrowed introgression, four genes were identified as transcription factors (Glyma03g28610, Glyma03g28630, Glyma03g28500, and Glyma03g28440; Wang et al., 2010).

Gene Expression Comparison between Clark and PI 547430

Previous studies in Arabidopsis have shown that AtbHLH038 is induced under iron-limiting conditions (Yuan et al., 2008). In our study, homologs of this gene (Glyma03g28610/28630) were also induced. Quantitative real-time (qRT)-PCR was used to confirm the expression pattern of these genes identified through RNA-Seq analysis. qRT-PCR confirmed the results of the RNA-Seq analysis. Both genes showed a greater than 2-fold induction under iron-insufficient conditions (Table III). While RNA-Seq data were not generated for PI 547430, qRT-PCR expression analysis was performed on the same genes in the inefficient NIL. A greater than 2-fold induction of Glyma03g28610 and Glyma03g28630 was observed in PI 547430 under iron stress conditions (Table III). Clark and PI 547430 did not show a significant difference in the expression levels of these genes under iron stress conditions.

In Arabidopsis, it has been shown that AtFIT binds AtbHLH038 as a heterodimer to activate AtFRO2 and AtIRT1 (Yuan et al., 2008). We used the available RNA from Clark and PI 547430 to investigate the expression patterns of these downstream genes. We identified the soybean homologs of the Arabidopsis genes using peptide similarity and found that Glyma07g07380 (GmFRO2) is homologous to AtFRO2, with 72.5% peptide similarity. Glyma07g34930 (GmIRT1) is homologous to AtIRT1, with 77.7% peptide similarity, and Glyma12g30240 (GmFIT) is homologous to FIT, with 62.1% peptide similarity. Using primers designed from these genes, qRT-PCR showed that Clark induced Glyma03g28610, Glyma03g28630, GmFIT, and the two downstream genes, GmFRO2 and GmIRT1, under iron stress (Table III). PI 547430 had significant induction of GmIRT1 yet failed to induce GmFIT or GmFRO2 under iron stress, even with high Glyma03g28610 and Glyma03g28630 expression (Table III). The expression levels of GmFIT, GmFRO2, and GmIRT1 in iron-stressed Clark were significantly higher than those in PI 547430.

Table II. Annotation of genes within narrowed introgression

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Differential Expression</th>
<th>PFAM Identifier</th>
<th>PFAM Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glyma03g28440</td>
<td>Leaves 1 and 2</td>
<td>PF00651</td>
<td>BTB/POZ domain</td>
</tr>
<tr>
<td>Glyma03g28450</td>
<td>Leaves 1 and 2</td>
<td>NA</td>
<td>Unknown function</td>
</tr>
<tr>
<td>Glyma03g28460</td>
<td>Leaves 1 and 2</td>
<td>NA</td>
<td>Unknown function</td>
</tr>
<tr>
<td>Glyma03g28470</td>
<td>Leaves 1 and 2</td>
<td>PF05834</td>
<td>Lycopene cyclase protein</td>
</tr>
<tr>
<td>Glyma03g28480</td>
<td>Leaves 1 and 2</td>
<td>PF03807</td>
<td>NADP oxidoreductase coenzyme F420 dependent</td>
</tr>
<tr>
<td>Glyma03g28490</td>
<td>Leaves 1 and 2</td>
<td>PF08241</td>
<td>Methyltransferase domain</td>
</tr>
<tr>
<td>Glyma03g28500</td>
<td>Leaf 1</td>
<td>PF07719</td>
<td>Tetraproricopeptide repeat</td>
</tr>
<tr>
<td>Glyma03g28510</td>
<td>Leaf 1</td>
<td>PF03871</td>
<td>RNA polymerase Rpb5, N-terminal domain</td>
</tr>
<tr>
<td>Glyma03g28520</td>
<td>NA</td>
<td>PF0462</td>
<td>Glutaredoxin</td>
</tr>
<tr>
<td>Glyma03g28530</td>
<td>NA</td>
<td>PF01053</td>
<td>Cys/Met metabolism PLP-dependent enzyme</td>
</tr>
<tr>
<td>Glyma03g28550</td>
<td>Leave 1 and 2</td>
<td>PF09713</td>
<td>Plant protein 1589 of unknown function (A_thal_3526)</td>
</tr>
<tr>
<td>Glyma03g28570</td>
<td>Leave 1 and 2</td>
<td>PF00072</td>
<td>Response regulator receiver domain</td>
</tr>
<tr>
<td>Glyma03g28580</td>
<td>Leaf 1</td>
<td>PF03088</td>
<td>Stricosidine synthase</td>
</tr>
<tr>
<td>Glyma03g28590</td>
<td>Leaf 1</td>
<td>PF03088</td>
<td>Stricosidine synthase</td>
</tr>
<tr>
<td>Glyma03g28600</td>
<td>NA</td>
<td>PF0702</td>
<td>Haloacid dehalogenase-like hydrolase</td>
</tr>
<tr>
<td>Glyma03g28610</td>
<td>Roots 1 and 2</td>
<td>PF00010</td>
<td>Helix-loop-helix DNA-binding domain</td>
</tr>
<tr>
<td>Glyma03g28630</td>
<td>Roots 1 and 2</td>
<td>PF00010</td>
<td>Helix-loop-helix DNA-binding domain</td>
</tr>
<tr>
<td>Glyma03g28640</td>
<td>NA</td>
<td>Unknown function</td>
<td></td>
</tr>
</tbody>
</table>
greater (3-, 5-, and 2.5-fold, respectively) than that of PI 547430 under iron stress.

Genomic DNA Sequencing of Differentially Expressed Transcription Factors

Glyma03g28610 and Glyma03g26830 were the only two transcription factors found to be differentially expressed within this region. These two genes were sequenced in Clark, Anoka, and T203 to determine if allelic differences could be identified through sequencing analysis. Allelic differences were identified between the iron-efficient Clark and the iron-inefficient T203 and Anoka. A 12-bp deletion was identified in Glyma03g28610 (Gm03:36552946.0.36552957), an in-frame four-amino-acid deletion (PELQ). Primers were designed to identify the presence or absence of the deletion and used to screen Harosoy, Lincoln, Richland, A7, Korean, PI 547698, and PI 547430. The 12-bp deletion was identified only in the iron-inefficient cultivars (Anoka, Korean, T203, PI 547698, and PI 547430). It was not detected in the iron-efficient genotypes (Clark, Harosoy, Lincoln, Richland, and A7; Fig. 5).

DISCUSSION

Whittling Donor Parent Introgressions

In this report, we have narrowed a large list of candidate genes derived from a classical genetic QTL study down to a manageable list of 18 genes. Two of the candidate genes are homologous to genes in Arabidopsis known to play a major role in iron acquisition. We used a variety of mapping techniques and additional backcrossing of NILs to narrow the region of interest on chromosome 3 down to a 250-kb region. The research reported herein shows that using the whole genome sequence of soybean, classical genetics, and genomic technologies, we were able to identify the genes underlying a QTL previously identified by Lin et al. (1997), nearly 15 years ago.

We investigated two sets of NILs sharing a common donor parent and identified regions of the genome introgressed from the donor parent into each of the recurrent parents. The results reported here using SSR markers to delineate the T203 introgressions in the PI 547430 NIL confirmed the introgressions using SNPs reported by Severin et al. (2010). The use of the common donor parent allowed for the overlay of the Clark and Harosoy introgression maps, significantly decreasing the proportion likely to contain the gene(s) involved in the iron stress response. Fortunately, the introgressed regions in each NIL shared only a small overlap, which allowed for further dissection of this region.

Early IDC QTL studies used the mapping population Anoka × A7 to identify an iron efficiency QTL on chromosome 3 that was responsible for more than 70% of the phenotypic variation (Lin et al., 1997). Since the late 1990s, the marker density on the soybean genetic map has been greatly improved. It was for this reason that the population used in this early QTL study was

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Annotation</th>
<th>Comparison</th>
<th>DE in RNA-Seq Set 1 (Clark)</th>
<th>DE in RNA-Seq Set 2 (Clark)</th>
<th>DE in qRT (Average)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glyma03g28610</td>
<td>BHLH038</td>
<td>CRS versus CRD</td>
<td>3.71a</td>
<td>4.47a</td>
<td>2.99a</td>
</tr>
<tr>
<td>Glyma03g28630</td>
<td>BHLH038</td>
<td>CRS versus CRD</td>
<td>6.24a</td>
<td>4.11a</td>
<td>2.26a</td>
</tr>
<tr>
<td>Glyma12g30240</td>
<td>FIT</td>
<td>CRS versus CRD</td>
<td>1.38</td>
<td>3.61a</td>
<td>5.93a</td>
</tr>
<tr>
<td>Glyma07g34930</td>
<td>IRT</td>
<td>CRS versus CRD</td>
<td>1.67</td>
<td>3.54a</td>
<td>3.97a</td>
</tr>
<tr>
<td>Glyma03g28610</td>
<td>BHLH038</td>
<td>ICRS versus ICRD</td>
<td>–</td>
<td>–</td>
<td>3.28a</td>
</tr>
<tr>
<td>Glyma03g28630</td>
<td>BHLH038</td>
<td>ICRS versus ICRD</td>
<td>–</td>
<td>–</td>
<td>3.09a</td>
</tr>
<tr>
<td>Glyma12g30240</td>
<td>FIT</td>
<td>ICRS versus ICRD</td>
<td>–</td>
<td>–</td>
<td>0.77</td>
</tr>
<tr>
<td>Glyma07g07380</td>
<td>FRO2</td>
<td>ICRS versus ICRD</td>
<td>–</td>
<td>–</td>
<td>0.70</td>
</tr>
<tr>
<td>Glyma07g34930</td>
<td>IRT</td>
<td>ICRS versus ICRD</td>
<td>–</td>
<td>–</td>
<td>2.11a</td>
</tr>
</tbody>
</table>

aGreater than 2-fold expression change.

Figure 5. PCR amplification of Glyma03g28610 with primers flanking the 12-bp deletion. Due to the high similarity between Glyma03g28610 and Glyma03g26830, the deletion is indicated by the double banding at 143 and 155 bp. The deletion is only present in the inefficient lines.

Table III. RNA-Seq and qRT-PCR expression data

Gene expression analysis of the known strategy I response genes using RNA-Seq and qRT-PCR is shown. Expression is reported by fold change between sufficient and insufficient iron conditions. CRS, Clark root sufficient; ICRS, PI 547430 root sufficient; CRD, Clark root deficient; ICRD, PI 547430 root deficient. DE, Differential expression.
rescreened with the approximately 900 newly available SSR markers (Soybase.org). Using the original phenotypic data, the major QTL on chromosome 3 was reconfirmed in our study and further refined. Lin et al. (1997) hypothesized that the QTL on chromosome 3 was the result of a single major gene within the QTL responsible for their phenotypic observations. Our findings with the NIL mapping in both the Harosoy and Clark backgrounds support their hypothesis and reconfirm the single major QTL in the Anoka × A7 mapping.

To identify the gene(s) responsible for iron inefficiency within the QTL, it was necessary to narrow the list of candidate genes. Previous studies have shown that additional backcrosses in NIL populations allow for narrowing the introgression boundaries (Nichols et al., 2006). The successes seen with this approach led us to develop the seventh backcross of PI 547430 to the recurrent parent Clark during summer 2009. We identified two distinct phenotypic classes: those containing the PI 547430 allele for BARCSOYSSR_03_1110 that displayed IDC symptoms similar to that of PI 547430, and those with the Clark allele for BARCSOYSSR_03_1110 that were healthy. From this, we were able to narrow the gene family further to a manageable 250-kb region, significantly reducing the number of candidate genes from more than 800 down to a mere 18. The 250-kb region colocates with the Anoka × A7 QTL on the centromeric side of the introgression overlap, indicating that Anoka × A7 and the two NILs may share common alleles for a gene or genes facilitating the IDC response.

Candidate Gene Expression

O’Rourke et al. (2009) performed microarray studies using the Clark/PI 547430 NIL and hypothesized that regulatory elements within the known iron QTL were responsible for gene expression changes of known iron genes located outside the QTL region. Our study supports that hypothesis. Four of the 18 genes identified within the narrowed introgression were annotated as transcription factors: Glyma03g28440, Glyma03g28500, Glyma03g28610, and Glyma03g28630 (Wang et al., 2010). Two of these genes, Glyma03g28610 and Glyma03g26830, are interesting as they are homologs of Arabidopsis AtbHLH038. It has been shown that AtbHLH038 and AtbHLH039 are closely related, functionally redundant transcription factors that each can form a heterodimer with AtFIT to enhance expression of the downstream iron acquisition genes AtFRO2 and AtIRT1 (Yuan et al., 2008). Polymorphisms in these two transcription factors likely underlie the QTL in question. Of the 18 genes identified within the narrowed region of interest, Glyma03g28610 and Glyma03g26830 were induced upon iron stress and were the only two genes found to be significantly differentially expressed in the roots and the only two transcription factors differentially expressed in either the roots or the leaves. qRT-PCR confirmed the expression trends of Glyma03g28610 and Glyma03g26830 in addition to the soybean homologs of AtFIT, AtFRO2, and AtIRT1. LeFER, the AtFIT homolog in tomato, is not induced upon iron stress (Ling et al., 2002), whereas AtFIT is induced upon iron stress along with AtbHLH038/039, which initiate the expression of AtFRO2 and AtIRT (Yuan et al., 2008). We have shown that FIT expression in soybean corresponds with the induction observed in Arabidopsis and not tomato, as soybean induces GmFIT expression upon iron stress.

Transcriptome analysis was not conducted on PI 54730; however, by using qRT-PCR, we were able to investigate the inefficient NIL’s response to iron stress at the gene expression level. PI 547430 showed no significant change in GmFIT or GmFRO2 expression between iron-sufficient and iron-insufficient conditions, although GmIRT1 was still induced greater than 2-fold but significantly less than that of Clark. AtIRT1 is typically thought to have coordinate expression with AtFRO2 (Connolly et al., 2003); however, our results showed a significant difference in the coordinated induction of GmFRO2 and GmIRT in PI 547430 under iron-limiting conditions. The lack of GmFRO2 induction is quite significant, since the reduction of Fe$^{3+}$ to Fe$^{2+}$ at the rhizosphere has been shown to be a rate-limiting step for iron uptake (Grusak et al., 1990; Connolly et al., 2003).

Although not analyzed in depth in this study, six genes were differentially expressed in Clark leaves. Of these genes, Glyma03g28570 stands out. It is annotated in the Kyoto Encyclopedia of Genes and Genomes database (Kanehisa and Goto, 2000) as a member of the two-component response regulator ARR-A family. To et al. (2004) identified the type A Arabidopsis response regulators as having a negative role in the regulation of cytokinins. Recently, Séguela et al. (2008) identified for the first time that cytokinins were involved in the regulation of AtFIT, AtFRO2, and AtIRT1. Potentially, Glyma03g28570 is another one of the underlying genes behind the significant effects of this iron efficiency QTL.

Interestingly, in the Anoka × A7 mapping population, the iron-inefficient parent Anoka was derived from the cross (Lincoln × Richland) × Korean. Clark, the iron-efficient recurrent parent used in the earlier study, contains the K-ras5A allele, which contains a mutation in the LeFER gene (Yuan et al., 2008). LeFER is a key regulator of Fe$^{2+}$ absorption, and its absence likely confers increased iron efficiency on Clark.

To identify the gene(s) responsible for iron inefficiency within the QTL, it was necessary to narrow the list of candidate genes. Previous studies have shown that additional backcrosses in NIL populations allow for narrowing the introgression boundaries (Nichols et al., 2006). The successes seen with this approach led us to develop the seventh backcross of PI 547430 to the recurrent parent Clark during summer 2009. We identified two distinct phenotypic classes: those containing the PI 547430 allele for BARCSOYSSR_03_1110 that displayed IDC symptoms similar to that of PI 547430, and those with the Clark allele for BARCSOYSSR_03_1110 that were healthy. From this, we were able to narrow the gene family further to a manageable 250-kb region, significantly reducing the number of candidate genes from more than 800 down to a mere 18. The 250-kb region colocates with the Anoka × A7 QTL on the centromeric side of the introgression overlap, indicating that Anoka × A7 and the two NILs may share common alleles for a gene or genes facilitating the IDC response.

Figure 6. Cartoon of the FIT/bHLH038 heterodimer bound to DNA. The protein model depicts the interaction of the strategy I response transcription factor FIT (yellow) with its heterodimer bHLH038, Glyma03g28610 (blue). The 12-bp deletion is indicated in red.
NIL mapping, was also derived from the Lincoln2 × Richland cross. From this, we would hypothesize that Korean donated the iron-inefficient allele(s) in Anoka. We mapped two distinct blocks of SSRs containing alleles derived from Korean on chromosome 3 that corresponded with the overlapped introgression identified by both of the NILs (data not shown). The identification of the same genomic region in the NIL and QTL mapping studies, along with the 12-bp deletion in Anoka and Korean, strongly indicate Glyma03g28610 or another gene or genes within this region as having a major effect on the iron efficiency of soybean. Furthermore, if the dimerization capabilities of this AtbHLH038 homolog are in fact hindered by the 12-bp deletion, it could explain why plants harboring this mutation are not as effective at surviving in low-iron environments.

Mutant Identification

The lack of a significant gene expression change for Glyma03g28610 and Glyma03g28630 between Clark and PI 547430 under iron stress conditions indicates that, if these genes are the underlying cause of the QTL, then the difference between the lines may be due to gene function. We identified two alleles of Glyma03g28610 that had 100% correlation with iron efficiency in this study. The iron-inefficient cultivars Anoka, T203, PI 547430, PI 547698, and PI 548360 all shared a 12-bp deletion within the second exon of this gene. The deletion is an in-frame four-amino-acid deletion that does not appear to hinder the accumulation of the gene’s transcript. According to the National Center for Biotechnology Information’s Conserved Domain Database, the deletion spans one of the 14 predicted dimerization sites and shifts two others (Fig. 6). This deletion could potentially alter Glyma03g28610’s ability to bind as a heterodimer with GmFIT or to alter the confirmation of the protein so that it is no longer able to bind the e-box within the promoter regions of GmFRO2 and GmIRT1.

In the iron-inefficient lines with the deletion, we hypothesize that Glyma03g28610 acts as a competitive inhibitor of the functional Glyma03g28630/FIT complex. A similar phenomenon has recently been shown in Arabidopsis. Hong et al. (2011), while studying floral architecture and leaf development, were able to show that truncated zinc finger transcription factors form nonfunctional heterodimers, which in turn act as a level of transcriptional regulation. This could explain why GmFRO2 is not induced in PI 547430 upon iron stress. It has been hypothesized that there is an additional regulatory step between AtFIT and AtIRT1 (Colangelo and Guerinot, 2004). This unknown factor could be the cause of the GmIRT1 expression in PI 547430. In addition, it has been shown that AtFIT regulates AtFRO2 at the level of transcription and IRT1 at the level of protein accumulation (Colangelo and Guerinot, 2004). It was beyond the scope of this study to look at protein accumulation. Thus, even though we observed an induction of GmIRT1 mRNA, we cannot claim that there is an increase of GmIRT1 protein in the soybean roots. PI 547430 lacked the ability to induce GmFIT under iron-limiting conditions, which indicates that there may be further unknown factors within the iron response pathway capable of inducing GmIRT1. Additionally, we can only state that the 12-bp deletion seems to be a causative factor in the parental genotypes studied here. A preliminary survey of a range of iron-efficient and iron-inefficient soybeans using the 12-bp deletion as a marker indicated that other factors also contribute to iron inefficiency in soybean.

CONCLUSION

In this report, we show that mapping sites of introgressions in multiple NILs sharing a common donor parent is a viable method for the identification of genomic regions controlling phenotype. Furthermore, through the creation of subNILs and gene expression analysis, we were able to identify two candidate genes within the NIL introgression on chromosome 3 that colocalize with a known iron-efficient QTL in the Anoka × A7 population. Finally, we identified a 12-bp deletion in one of the candidate genes that is shared in the inefficient lines in both NILs and the inefficient parent, Anoka, in the Anoka × A7 population.

MATERIALS AND METHODS

Introgression Mapping

Soybean (Glycine max) lines used in this study were derived from backcrossing the two efficient recurrent parents Clark and Harosoy crossed with the common donor parent T203 (Bernard, 1975). Mapping of the T203 (donor) introgression into both PI 547430 (Clark background) and PI 547698 (Harosoy background) utilized 869 and 864 SSR markers, respectively. Markers were used to amplify DNA from the donor, from recurrent parents, and from the isolines. The PCR product was visualized on a 6% polyacrylamide gel. Screened markers were categorized into three classes: not informative, informative between the parents, and introgressed DNA. The introgressed class was defined by the isolate’s banding pattern matching that of the donor parent. All markers had a known location on the whole genome sequence (Glyma1.01; Soybase.org) and thus identified regions of introgression into the NIL.

QTL Mapping in the Anoka × A7 Population

Plant material for QTL mapping originated from 92 F2:4 lines from the Anoka × A7 population, which was previously used to map IDC QTLs (Lin et al., 1997). The genetic linkage map was created by screening the population with 916 known SSR markers (Soybase.org) and an additional 303 untested BARCSOYSSR markers (Song et al., 2010). The marker scores for each line were then imported into Mapmaker 3.0 (Lander et al.,1987). A LOD threshold of 4.0 was used in a three-point linkage analysis to construct the map.

Phenotypic data obtained by Lin et al. (1997) were used in the research reported here. Linkage maps were imported into MAPQTL6 (Van Ooijen, 2009), and QTLs were mapped using interval mapping and multiple QTL mapping (Lander and Botstein, 1989; Jansen, 1993, 1994). A significant experiment-wise LOD threshold of 3.8 was determined by 1,000 permutations of the data.

Backcross Advancement of PI 547430

To further increase the chances of narrowing the introgressed region within the NIL (Clark background), an additional backcross generation was performed in these lines during the summer of 2009. Seventy-five Clark and 75 PI
547430 seeds were planted on May 19, 2009, at Brunner Farm (Ames, IA). Two hundred crosses were performed that produced 156 seeds, which were planted during the 2009/2010 winter season in the greenhouse at Ames, Iowa. The BC$_2$ plants were screened with four SSR markers to confirm heterozygosity within the introgressed region. Forty seeds from each confirmed cross were then planted at Brunner Farm on May 27, 2010. Approximately 2,600+ plants were screened with Satt387, Satt521, Sat$_{091}$, and Sat$_{295}$ to identify recombinations within the introgressed region. Of these, 480 plants with recombinations within the introgression region were selected for genotyping with four additional SSR markers (Sat$_{236}$, Satt549, Satt339, and Sat$_{304}$) to better delineate the recombination intervals.

Hydroponic Phenotyping

In the winter of 2010, seeds of each line were sown on germination paper for 7 d before seedling transfer to a custom plant holder suspended over a 10-L, foil-wrapped plastic bucket. Two replicates containing eight plants in each recombinant class were grown under iron stress conditions ([50 μM Fe(NO$_3$)$_3$·9H$_2$O], following the procedure outlined by O’Rourke et al. (2007). Each 10-L bucket contained 2 mM MgSO$_4$·7H$_2$O, 2 mM Mg(NO$_3$)$_2$·6H$_2$O, 2.5 mM KNO$_3$, 1 mM CaCl$_2$·2H$_2$O, 4 mM Ca(NO$_3$)$_2$·4H$_2$O, 0.02 mM KH$_2$PO$_4$, 542.5 μM H$_3$BO$_3$, 217 μM KOH, 20 μM MnCl$_2$·2H$_2$O, 50 μM ZnSO$_4$·7H$_2$O, 50 μM CuSO$_4$·5H$_2$O, 0.2 μM Na$_2$MoO$_4$·2H$_2$O, 1.0 μM CoSO$_4$·7H$_2$O, 1.0 μM NiSO$_4$·6H$_2$O, and 10 μM H$_2$BO$_3$, in accordance with research by R.L. Chaney and P.F. Bell (unpublished data) and O’Rourke et al. (2007). A 3% CO$_2$-air mixture was used to maintain a pH level of 7.8. Each day, a stock solution containing 30 mM K$_2$HPO$_4$, 222 mM KH$_2$PO$_4$, and 0.179 mM H$_2$BO$_3$, was added to the nutrient solution system to replace nutrients utilized by the growing plants. The greenhouse photoperiod was held constant at 14 h with a 10-h dark period for the entirety of the experiment. Plants were phenotyped 15 d after transfer into hydroponics at the V3 stage. The 1 to 5 visual IDC score identification was performed as outlined by Cianzio et al. (1979). In addition to visual score, SPAD readings were also recorded with a Minolta SPAD-502 chlorophyll meter immediately following visual score determination.

RNA Sampling of Clark and PI 547430 Tissues following Iron Shock (Root and Leaf)

Plants of Clark and PI 547430 were grown using a similar greenhouse hydroponic system with sufficient amounts of iron for 14 d according to the methods described above. After this period, plant roots were rinsed six times to remove iron and other nutrient traces and transferred to either iron-sufficient [100 μM Fe(NO$_3$)$_3$·9H$_2$O] or iron-insufficient [30 μM Fe(NO$_3$)$_3$·9H$_2$O] hydroponic conditions. After 24 h, foliar and root tissue was collected and flash frozen for storage. RNA was later extracted and purified using the RNeasy mini kit (Qiagen). The isolated RNA was then treated with DNase to remove final traces of DNA and stored at −80°C until sequenced.

Illumina Next-Generation Sequencing of RNA

Four libraries were sequenced using Genome Analyzer II (Illumina). The libraries consisted of two iron-sufficient and two iron-insufficient biological replicates each containing three pooled replicates. Sequencing was performed using established methods at the National Center for Genome Resources. Severin et al. (2010) have outlined the techniques of RNA-Seq sequencing. In brief, RNA with a poly(A) tail was isolated and size selected for approximately 500-bp fragments. PCR-amplified products of the size-selected library were PCR-amplified with test primers to that of the reference gene (ELF-1B). The technical and biological replicates were then averaged and log base 2 transformed.

Analysis of Next-Generation Sequencing Data

The 36-bp reads were aligned to the soybean genome transcript map (Gmax1.09; Schmutz et al., 2010) using Bowtie/Tophat (Trapnell et al., 2009). Alignment was limited to a seed length of 28 bases with a mismatch limit of two. The alignments were further processed using Cufflinks/Cuffdiff version 1.0.2 (Roberts et al., 2011). Transcript abundances were determined in Cufflinks using the fragments per kilobase of exon per million fragments mapped, similar to the method described by Mortazavi et al. (2008). Normalization of samples was performed using upper quantile normalization in order to better achieve estimates of lowly expressed genes (Bullard et al., 2010). Cuffdiff was used to identify gene expression differences between sufficient and insufficient iron conditions (Roberts et al., 2011).

Candidate Genes Underlying an Iron Efficiency QTL

differential gene expression was determined using a false discovery rate of 0.05 with a minimum of 30 alignments at a locus for valid significance testing and a significant P value cutoff of 0.05.

Gene Expression Analysis through qRT-PCR

Primers were designed from protein-coding sequences (Phytozome.net) using Primer3 as a primer design tool (Rozen and Skaletsky, 2000). Primers were designed to amplify approximately 100-bp fragments ideally spanning the intron/exon border to differentiate RNA- and DNA-amplified product. RNA amplification was performed using the SuperScript III Platinum SYBR Green One-Step qRT-PCR amplification kit (Invitrogen). Reactions were performed on a Stratagene Mx3000p real-time PCR machine (Agilent Technologies). Reactions were run for 44 cycles at a temperature of 60°C. A standard curve was created using 2.5-, 12.5-, 25-, 100-, and 150-ng RNA concentrations along with a no-template control and a no-reverse transcriptase well for each primer pair.

ELF-1B was chosen as the reference gene for this study based on a previous report by Wang et al. (2009). For each plate, one test primer and one reference gene primer were screened with two technical replicates for each of the three biological replicates. This design allowed for comparison between the quantity values (Rieu and Powers, 2009). qRT-PCR was used for analysis only if the dissociation curve showed one peak and the no-reverse transcriptase and no-template reactions were greater than five cycles away from the other data points. The standard curve was accepted if the $r^2$ value was greater than 0.985.

The relative quantities were determined by normalizing the values for the test primers to that of the reference gene (ELF-1B). The technical and biological replicates were then averaged and log base 2 transformed.

Genomic DNA Sequencing

DNA was extracted from Clark, PI 548533, Harosoy (PI 548877), T203 (PI 54619), PI 547430, PI 547468, Anoka (PI 548508), and A7 (PI 596526) using the 96-well hexadecyltrimethylammonium bromide (CTAB) method outlined by Dietrich et al. (2002). Primers were designed across the predicted gene calls starting approximately 500 bp upstream of the transcription start site and continuing approximately 500 bp past the transcription stop. Individual primers were designed to amplify approximately 750-bp fragments. The sequencing methods used are outlined by Severin et al. (2010). Cleanup of the PCR results utilized exonuclease 1 and shrimp alkaline phosphatase. Sequencing was performed using the BigDye version 3.1 sequencing kit on a 50-cm, 96-capillary 3730XL DNA analyzer (Applied Biosystems). The sequences from all eight genotypes were then aligned using Sequenser version 4.9 (Gene Codes Corp.). The alignments were used to identify SNPs between the sequenced lines.

ACKNOWLEDGMENTS

We thank Rebecca Nolan, Johanna Dobbs, Sarah Atwood, and Jenna Woody for their valuable discussions and support through this work.

Received October 24, 2011; accepted January 29, 2012; published February 7, 2012.

LITERATURE CITED


