Genome-wide RNA-seq Analysis of Differentially Expressed Transcription Factor Genes Against Bacterial Leaf Pustule in Soybean

Kil Hyun Kim, Yang Jae Kang, Sangrea Shim, Min-Jung Seo, Seong-Bum Baek, Jeom-Ho Lee, Sang Koo Park, Tae Hwan Jun, Jung-Kyung Moon, Suk-Ha Lee, Chang-Hwan Park

INTRODUCTION

Transcription factors (TFs) bind to specific DNA sequences and interact with other transcriptional regulators, such as co-activators, co-repressors and chromatin remodeling/modifying proteins, leading to the activation or repression of the expression of target genes. TFs influence many diverse biological processes, such as cellular division, development, and differentiation. They also control metabolic and physiological balance after environmental stimuli (Riechmann and Ratcliffe 2000; Czechowski et al. 2004). In particular, the specific interactions between TFs and cis-regulatory elements (CREs) play the important roles in plant responses to a range of abiotic and biotic stresses (Yamaguchi-Shinozaki and Shinozaki, 2005; Priest et al. 2009).

After the completion of genome sequence of Arabidopsis thaliana, genes coding for TFs have been identified from this model plant (Riechmann et al. 2000). Iida et al. (2005) have identified 1,968 proteins as transcription factors, accounting for 7.4% of all Arabidopsis genes. With known DNA-binding domain sequences, putative TF gene sets have also been predicted in legumes including Lotus japonicus, Medicago truncatula and Glycine max (Udvardi
et al. 2007; Libault et al. 2009). Recent completion of soybean genome sequence has revealed that soybean contains 64 TF families with a total of 5,671 TF genes, representing 12.2% of 46,430 predicted soybean protein-coding loci (Schmutz et al. 2010). Based on this information, SoybeanTFDB (http://soybeanTFDB.psc.riken.jp) and SoyDB (http://casp.met.missouri.edu/soydb) have been developed independently (Mochida et al. 2009; Wang et al. 2010b). Although most of these TF gene sequences are available, a relatively limited number of TF genes have been genetically characterized in soybean (Zhou et al. 2008; Zhang et al. 2009). Even in legume model plants L. japonica and M. truncatula, the functional roles of less than 1% of TF genes have been reported. Expediting the genetic characterization of the functions of putative TF genes requires genome-wide analysis of TF expression profiling in diverse plant tissues, developmental stages and environmental conditions (Udvardi et al. 2007; Libault et al. 2010).

Bacterial leaf pustule (BLP) is a serious disease caused by Xanthomonas axonopodis pv. glycines (Xag). Typical symptoms of BLP are small yellow to brown lesions with raised pustules in the center, resulting in yield loss by premature defoliation (Hartwig and Johnson 1953; Groth and Braun 1986). Although rxp gene is a known BLP resistance gene, its sequence information has not been reported (Palmer et al. 1992; Narvel et al. 2001). To overcome this limitation, near isogenic lines (NILs) differing in the presence or absence of the target gene have been developed by multiple backcross and foreground as well as background selections using simple sequence repeat (SSR) markers (Kim et al. 2008).

In our previous study, RNA-Seq analysis has been performed to study plant responses to Xag using BLP-resistant and resistant NILs (Kim et al. 2011). Pathogen associated molecular patterns (PAMP), damage-associated molecular pattern (DAMP) receptors, and genes induced by these two receptors are up-regulated in BLP resistant NIL. In particular, MYC2, WRKY33, and Jasmonate ZIM-motif, the core of jasmonic acid (JA)-signaling components, are highly expressed in BLP-resistant NIL, suggesting that a lot of TFs might be related to BLP resistance in soybean. The objectives of this study were to identify differentially expressed TF genes between BLP-susceptible and BLP-resistant NILs at three time points (0, 6 and 12 hours) after Xag infection) using SoybeanTFDB and SoyDB (Mochida et al. 2009; Wang et al. 2010b) and to determine putative BLP resistance genes mediated by the recruitment of TFs to cis-regulatory elements (CREs).

**MATERIALS AND METHODS**

**Plant Materials and Xag Inoculation**

NILs were constructed by three cycles of repeated backcrossing of BLP-resistant SS2-2 as donor parent with BLP-susceptible Taekwangkong as recurrent parent (Kim et al. 2008). Soybean plants were cultivated in growth chambers at 28°C with 12 hr illumination. The Xanthomonas axonopodis pv. glycines (Xag) strain 8ra was cultured on peptone sucrose agar medium at 28°C for 48 hr (Oh et al. 1999) supplemented with 0.1 ppm rifampicin. The bacterial culture was diluted to obtain 1 × 10^8 colony forming units (cfu) per milliliter with 10 mM MgCl_2 with an optical density of 0.3–0.5 at 600 nm. This suspension was sprayed onto first fully expanded trifoliate leaves of one month-old soybean plants using an atomizer. All inoculated plants were grown in almost 100% relative humidity.

**Illumina Sample Preparation and Clustering**

Total RNAs were extracted from plants using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Poly-A containing mRNA was isolated from total RNA using oligo-dT magnetic beads. After mRNA was fragmented into small pieces, they were used as templates for cDNA synthesis. Sheared cDNAs were end repaired and ligated to Illumina adaptors following manufacturer’s protocols using mRNA-Seq Sample Preparation kit (Cat. RS-930-1001, Illumina, Inc., San Diego, CA, USA). cDNA fragments were loaded onto Illumina flow cell containing eight lanes, including an internal control. One sample of 6 hai of BLP-resistant NIL was loaded into two lanes to test mechanical replication. Using single-read cluster generation kit v2 on the cluster station (Illumina, Inc.), DNA templates were used for
cluster generation and cluster amplification. 76 bp reads were collected from Illumina Genome Analyzer II using sequencing-by-synthesis technology. Image data was acquired from the sequencing run by the Illumina platform.

**Statistical Analysis**

76 bp reads were aligned to contigs of the 8x soybean genome sequence assembly (Glyma1.01 genome assembly) using Bowtie (http://bowtie-bio.sourceforge.net; Langmead et al. 2009). TopHat (http://tophat.cbcb.umd.edu) was used to identify novel transcripts from the location of regions covered and to estimate the abundance of transcripts from their depths of coverage in the mapping (Trapnell et al. 2009). DEGseq (http://bioinfo.autsinghua.edu.cn/software/degseq) was used to identify differentially expressed genes between BLP-susceptible and BLP-resistant NILs (Wang et al. 2010a). Mapped read counts were converted into reads per kilobase pair of transcript per million mapped reads (RPKM) (Motazabi et al. 2008). After filtering with more than two-fold differences and P-value less than 0.001 after log<sub>2</sub> transformed, differentially expressed TF genes at each time periods (0, 6, and 12 hr) were obtained from SoyDB (Wang et al. 2010b).

**Discovery of cis-Regulatory Motifs in Promoter Regions of TF Genes**

CREs bound by transcription factors are known to regulate the expression of corresponding genes. Short sequences of these CREs have been collected from several previous studies. There are various CREs in the promoter regions of stress inducible genes (Yamaguchi-Shinozaki and Shinozaki 2005). Their cis-regulatory motifs were summarized in Table S2. To determine which genes might be influenced by TFs, we collected 2-kb upstream sequences from the transcription start site and searched for cis-motifs using Python script.

**RESULTS**

**Identification of differentially expressed TF genes and cis-regulatory elements for BLP resistance between BLP-susceptible and BLP-resistant NILs**

A total of 15,678 genes were obtained from RNA-Seq datasets after *Xag* inoculation (Kim et al. 2011). After being log<sub>2</sub> transformation followed by filtering of more than two-fold differences, a total of 2,415 differentially expressed genes were obtained. A total of 65,781 predicted genes were collected from Phytozone (Kim et al. 2011). After being filtered for more than two-fold differences and P-value less than 0.001, 46,367 mapped genes were obtained. After being log<sub>2</sub> transformation, 15,678 genes were obtained with a RPKM value. After filtering with more than two-fold differences and P-value less than 0.001, 2,415 significant genes were obtained. After being filtered with more than two-fold differences, 351 TF genes were obtained. After being filtered with more than two-fold differences and P-value less than 0.001, 1,092 downstream genes were obtained. Understanding the role of TF genes in plant defense mechanism is critical.

![Flow chart to understand the roles of transcription factor genes in soybean defense mechanism.](image-url)
expressed genes were found (Fig. 1, yellow box). To identify differentially expressed TF genes between BLP-susceptible and BLP-resistant NILs at each time point (0, 6, and 12 hai), 5,683 TF genes belonging to 63 annotated TF families in SoyDB (http://casp.rnet.missouri.edu/soydb/) were surveyed. A total of 351 TF genes encoding 33 different TF families were identified from these 2,415 genes (Fig. 1, orange box and Table S1).

Since various TFs can interact with cis-regulatory elements (CREs) to control gene expression, we tried to identify TFs and their CREs involved in the regulation of BLP-responsive transcription. Among well characterized CREs involved in the regulation of stress-responsive transcription, abscisic acid-responsive element (ABRE), G-box, MYB recognition site (MYBR), MYC recognition site (MYCR) and W-box were selected. More than three hits observed in the 2 kb upstream of transcription start site in the promoter region was used as cut-off (Fig. 1, red box). A total of 1,092 downstream genes containing ABRE, G-box, MYBR, MYCR, or W-box were identified (Supplementary Table 2). With 351 TF genes and 1,092 downstream genes, it was possible to study how plant immunity occurred via the interactions between TFs and CREs (Fig. 1, green box).

**Identification of TF families associated with BLP resistance**

To determine the distribution of BLP-responsive TF families after Xag treatment, top 10 TF families were compared to total TF families of SoyDB (Table 1). These TF families occupied almost 80% of the BLP NIL set. However, they only occupied about 50% in SoyDB. Of these BLP-responsive TF families, the proportions of AP2-EREBP, MYB, bHLH, WRKY, NAC, GRAS, C2C2 (Zn) CO-like and ZIM were higher than those in SoyDB except C2H2 (Zn) and TPR. MYB TF family (MYB and the MYB/HD-like) appeared to be the largest one in SoyDB, while AP2-EREBP was the largest one in BLP NIL. The proportions of AP2-EREBP, C2C2 (Zn) CO-like, WRKY, and ZIM TF families in BLP NILs were more than doubled compared to that in SoyDB (Table 1), suggesting that these 10 TF families play important roles in BLP resistance.

### Table 1. Distribution and number of differentially expressed transcription factor genes between BLP-susceptible and BLP-resistant NILs compared to SoyDB (http://casp.rnet.missouri.edu/soydb/).

<table>
<thead>
<tr>
<th>TF family</th>
<th>BLP-responsive TF gene</th>
<th>SoyDB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage (%)</td>
</tr>
<tr>
<td>AP2-EREBP</td>
<td>67</td>
<td>19.09</td>
</tr>
<tr>
<td>MYB/HD-like</td>
<td>50</td>
<td>14.25</td>
</tr>
<tr>
<td>bHLH</td>
<td>35</td>
<td>9.97</td>
</tr>
<tr>
<td>WRKY</td>
<td>31</td>
<td>8.83</td>
</tr>
<tr>
<td>NAC</td>
<td>24</td>
<td>6.84</td>
</tr>
<tr>
<td>GRAS</td>
<td>16</td>
<td>4.56</td>
</tr>
<tr>
<td>C2C2 (Zn) CO-like</td>
<td>14</td>
<td>3.99</td>
</tr>
<tr>
<td>C2H2 (Zn)</td>
<td>14</td>
<td>3.99</td>
</tr>
<tr>
<td>TPR</td>
<td>13</td>
<td>3.70</td>
</tr>
<tr>
<td>ZIM</td>
<td>13</td>
<td>3.70</td>
</tr>
<tr>
<td>Others</td>
<td>74</td>
<td>21.08</td>
</tr>
<tr>
<td>C2H2 (Zn) Dof</td>
<td>10</td>
<td>2.85</td>
</tr>
<tr>
<td>Homeodomain/HOMEOBOX</td>
<td>8</td>
<td>2.28</td>
</tr>
<tr>
<td>bZIP</td>
<td>6</td>
<td>1.71</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>351</td>
<td><strong>100.00</strong></td>
</tr>
</tbody>
</table>
Distribution of 351 differentially expressed TF genes at each time point

We surveyed the distributions of 351 up- or down-regulated TF genes in BLP-resistant NIL compared with BLP-susceptible NIL at each time course (0, 6, and 12 hr; Fig. 2). Of the 351 TF genes, 263 and 40 were up-regulated and down-regulated, respectively (Fig. 2A). The remaining 48 TF genes were up- or down-regulated in BLP-resistant NIL at each time point. Most TF genes were up-regulated in BLP-resistant NIL. Before Xag inoculation (0 hr), 238 TF genes were up-regulated in BLP-resistant NIL. However, 11 TF genes were down-regulated (Fig. 2B). At 6 hour after inoculation (hai), 18 TF genes were up-regulated, whereas 68 TF genes were down-regulated. At 12 hai, 90 TF genes were up-regulated, whereas 10 TF genes were down-regulated. On the contrary, most TF genes were down-regulated in BLP-susceptible NIL. 68 TF genes involved in AP2-EBEBP, bHLH, GRAS, HOMEOBOX, and MYB TF families were mainly up-regulated at 6hai (Fig. 2B and Table S1).

Our data revealed that most TF genes were highly expressed in BLP-resistant NIL at 0 hr. The top 10 major TF families that were up-regulated were listed in Table 1. After Xag inoculation, the number of differentially expressed BLP-responsive TF genes was dramatically reduced and down-regulation at 6 hai. However, most of them were up-regulated in BLP-resistant NIL at 12 hai.

**Fig. 2.** Number of differentially expressed transcription factor genes between BLP-susceptible NIL and BLP-resistant NIL 

[p < 0.001 and log2 (fold change)]. (A) The number inside the Venn diagram indicates the number of TF genes up-regulated or down-regulated. (B) Time-wise distribution of differentially expressed genes between BLP-susceptible and BLP-resistant NILs after Xag treatment. Horizontal bar represented time periods and vertical line represented the total number of differentially expressed TF genes at various time points.
Gene ontology (GO) analysis of differentially expressed TF genes using SoybeanTFDB

SoybeanTFDB (http://soybeantfdb.psc.riken.jp) has assessed the putative functions of soybean TF genes via comparative analysis with GO annotations of Arabidopsis with well-described GO terms of the biological process level (Mochida et al. 2009). Therefore, these assigned GO terms were applied to differentially expressed TF genes. After excluding GO terms such as ‘biological process’, ‘DNA-dependent’, and ‘regulation of transcription’, the distributions of top 19 most abundant GO terms in response to Xag treatment were analyzed (Fig. 3). Significant proportions of these differentially expressed TF genes were related to stress, hormone, and defense responses. The most highly representing GO terms were genes associated with chitins and salt stress. Plant hormone was mainly related to MYB TF family. In particular, salicylic acid was associated with MYB and WRKY TF families, while MYB and ZIM TF families were associated with JA. Interestingly, in defense response to bacterium (GO:0042742), all GO terms presenting WRKY and ZIM TF families were up-regulated in BLP-resistant NIL at 0 hr.

Downstream genes containing cis-regulatory elements

To understand the implication of cis-regulatory elements (CREs) in regulating BLP-responsive transcription, the following five CREs were selected: abscisic acid-responsive element (ABRE), G-box, MYB recognition sites (MYBR), MYC recognition sites (MYCR), and W-box (Table S2). The 2 kb upstream sequences of transcription start site were

Fig. 3. Distribution of GO term of biological processes in soybean TF genes after Xag treatment. Excluding regulation of transcription, DNA-dependent, and biological process, the top 19 most abundant terms were used to classify these TFs (SoybeanTFDB, http://soybeantfdb.psc.riken.jp).
surveyed for the 2,415 differentially expressed genes identified by RNA-Seq analysis. These downstream genes were selected for further analysis, if their promoter regions contained \( \geq 3 \) of the same type of CREs regardless of the type of CREs. As a result, a total of 1,092 downstream genes containing five different CREs were identified (Fig. 4 and Table S2).

Of these 1,092 downstream genes that were significantly expressed between BLP-susceptible and BLP-resistant NILs, 41, 278, 4, 7, and 905 genes contained ABRE, MYCR, MYBR, G-box and W-box, respectively. The most distinguishable feature was that out of 1,092 genes, a total of 905 genes had W-box known to be associated with biotic stress. Of the 1,092 genes, 139 contained more than two different types of CREs, whereas the remaining 953 had only one type of CRE. Our data suggested that genes containing W-box and MYCR might be related to BLP resistance.

**Chromosomal distribution of transcription factors and their downstream genes containing CREs with respect to BLP and disease-related QTLs**

Chromosomal locations of BLP-responsive TF genes in five TF families (AP2, bHLH, MYB, NAC, and WRKY) and their downstream genes containing CREs (ABRE, G-box, MYBR, MYCR and W-box) were plotted on a circular soybean genome map (Fig. 4). The five TF families responding to BLP were not located on all 20 soybean chromosomes. For example, Chromosome (Chr) 1 and Chr 2 did not have differentially expressed bZIP TF genes. Circos map showed that genes of TF families and their corresponding CREs were not co-localized. TF and CREs were not counterparts (Table S2). Two different CREs such as G-box and ABRE were represented in the promoter regions of downstream genes that had AP2-binding site. Most downstream genes contained W-box. These genes were distributed in all 20 soybean chromosomes. Disease-related QTLs were placed on the Circos map after the information was retrieved from SoyBase (http://soybase.org, Fig. 5). Disease-related QTLs were identified from all 20 soybean chromosomes except Chr 4. Especially, Chrs 16, 17, and 18 had many QTLs related to various pathogens. Circos map enabled us to look at genome-wide distribution of TF genes and their CREs compared to disease-related QTLs.

![Fig. 4. Significantly expressed downstream genes containing cis-regulatory elements (ABRE, G-box, MYBR, MYCR, and W-box) between BLP-susceptible and BLP-resistant NILs. The number inside the parentheses indicates the number of each cis-regulatory element.](image-url)
Fig. 5. Overview of the distribution of transcription factors, their downstream genes containing cis-regulatory elements, and disease related QTLs by soybean chromosome. This figure was drawn using circular genome data visualization software Circos (http://circos.ca/). TF families, AP2 (yellow-green), bHLH (grass-green), MYB (royal-blue), NAC (pink), and WRKY (brown), were positioned on bright grey layers. Grey layers showed their downstream genes containing specified CREs, such as ABRE (blue), MYBR (green), G-box (spring grass), MYCR (yellow) and W-box (orange). Additionally, a total of 296 disease-related QTL locations based on SoyBase were represented at the outermost layer besides chromosomes. The most inner layer visualized duplicated positions of the G. max genome, representing as ribbons after similar duplicated regions were grouped as bundles.
DISCUSSION

In previous study, a total of 2,415 differentially expressed genes were identified between BLP-susceptible and BLP-resistant NILs in response to Xag infection using RNA-Seq (Kim et al. 2011). Among them, we investigated 134 putative defense response genes in soybean based on BLASTP analysis of the defense response gene (GO:0006952). Most genes were related pathogen-associated molecular pattern (PAMP)-triggered immunity, effector-triggered immunity, pathogenesis-related genes, and transcription factors. Since many TFs and CREs were involved in disease resistance, we surveyed TFs and their downstream genes containing CREs in 2,415 differentially expressed genes identified between BLP-susceptible and BLP-resistant NILs. With SoyDB and five selected CREs, we tried to understand the role of BLP-responsive TF genes and their downstream genes in plant defense (Fig. 1).

A total of 351 differentially expressed TF genes within 33 annotated TF families were identified between BLP-susceptible and BLP-resistant NILs using SoyDB (Table 1). Top 10 major TF families accounting for approximately 80% of differentially expressed TFs had significant participation in transcriptional regulation after Xag treatment. These TF families might be related to BLP resistance. The percentages of AP2-EREBP, C2C2 (Zn) CO-like, WRKY, and ZIM TF families in the BLP-responsive TF genes were 2-fold higher than that in SoyDB. According to Guo et al. (2004), over-expression of the AP2/EREBP transcription factor osmotin promoter binding protein (OPBP1) enhanced bacterial disease resistance in tobacco. It activated the expression of the PR-1a and PR-5d genes, leading to enhance resistance to infection by Psedomonas syringae pv tabaci and Phytophthora parasitica var nicotianae pathogens. The WRKY TF family has been demonstrated to be involved in regulations of responses to various abiotic and biotic stresses (Eulgem and Sommssich 2007). The Arabidopsis plants over-expressing AtWRKY28 and AtWRKY75 enhanced resistance to oxalic acid and Sclerotinia sclerotiorum. Over-expression of these genes induced oxidative burst in host plants and consequently triggered plant resistance to infection by pathogen (Chen et al. 2013). Our results are consistent with these previous studies of OPBP1 (Glyma-18g48730), AtWRKY28 (Glyma05g25770), and AtWRKY75 (Glyma03g05880).

Furthermore, AP2-EREBP, C2C2 (Zn) CO-like, WRKY, and ZIM TF families have significantly over-represented among the differentially expressed genes under dehydration and salt stress (Belamkar et al. 2014).

As shown in Fig. 2, a total of 238 TF genes were up-regulated in BLP-resistant NIL at 0 hr, whereas only 11 TF genes were down-regulated. At 6 hr after Xag infection, more TF genes were identified as down-regulated. However, the number of up-regulated TF genes was increased again at 12 hr after Xag infection. These results suggest that BLP-resistant plants can recognize bacterial infection and transduce their perceived signals to the downstream pathways very quickly to defend against Xag. Bieri et al. (2004) have also reported a rapid turnover of resistance protein to help steady state plant against stress.

Five TF families (AP2-EREBP, bHLH, MYB, NAC and WRKY) were distributed throughout the genome. A total of 905 W-box containing downstream genes were positioned on every soybean chromosome (Fig. 4). Soybean genome undergoes whole genome duplication (Van et al. 2008). Gene duplications of TFs in response to Xag were also observed in this study (represented as gray ribbons in Fig. 4). For example, all five TF genes were located at the region near QTL for Sclero on Chr 5. However, their duplicated regions on Chr 8 did not have all five TF genes. After QTLs related to disease resistance were placed on soybean chromosomes, many disease-related QTLs in response to various pathogens were positioned on Chrs 16, 17, and 18.

Recessive gene rxp that conditions resistance to BLP (Palmer et al. 1992; Narvel et al. 2001) is located between Satt486 (Gm17:7,124,028..7,124,278) and Satt372 (Gm17: 7,811,252..7,811,508). Close to this region, two differentially expressed genes, Glyma17g09850 (TPR TF) and Glyma-17g10040 (PF07887: calmodulin binding protein-like) were identified between BLP-resistant and BLP-susceptible NILs. The following four candidate TF genes could be selected to be associated with BLP resistance if the expanded region was considered: Glyma17g07640 (bHLH), Glyma17g08380 (MYB), Glyma17g10820 (MYB) and Glyma17g10920 (C2C2 (Zn) Dof). Further expression analysis of these candidate TF genes on Chr. 17 could
provide useful information on defense mechanism against BLP.

In conclusion, 351 differentially expressed TF genes were identified between BLP-susceptible and BLP-resistant NILs after Xag inoculation. Most of them belonged to top 10 major TF families. Among them, five TF families (AP2-EREBP, bHLH, MYB, NAC, and WRKY) were highly up-regulated in BLP-resistant NIL. Most downstream genes containing MYCR and W-box could be related to BLP resistance in soybean. These efforts will enable us to understand the signal pathway involved in bacterial pathogenesis. It will contribute to the improvement of soybean by genetic engineering programs to grow in various stress conditions.

ACKNOWLEDGEMENTS

This study was principally supported by a post-doctoral fellowship (Project No. PJ009330) funded by the National Institute of Crop Science Research Program, Rural Development Administration (RDA), Republic of Korea.

REFERENCES


Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran LSP. 2009. In silico analysis of
transcription factor repertoire and prediction of stress responsive transcription factors in soybean. DNA Res. 16: 353-369.


