MPMI Vol. 24, No. 12, 2011, pp. 1530–1539. doi:10.1094/MPMI-05-11-0106. © 2011 The American Phytopathological Society *e*-Xtra*

Digital Gene Expression Profiling of the *Phytophthora sojae* Transcriptome

Wenwu Ye,[1] Xiaoli Wang,[1] Kai Tao,[1] Yuping Lu,[1] Tingting Dai,[1] Suomeng Dong,[1] Daolong Dou,[1] Mark Gijzen,[2] and Yuanchao Wang[1]

[1]Department of Plant Pathology, Nanjing Agricultural University, Nanjing 210095, China; [2]Agriculture and Agri-Food Canada, 1391 Sandford Street, London, Ontario N5V 4T3, Canada

Submitted 13 May 2011. Accepted 11 August 2011.

The transcriptome of the oomycete plant pathogen *Phytophthora sojae* was profiled at ten different developmental and infection stages based on a 3′-tag digital gene-expression protocol. More than 90 million clean sequence tags were generated and compared with the *P. sojae* genome and its 19,027 predicted genes. A total of 14,969 genes were detected, of which 10,044 were deemed reliable because they mapped to unambiguous tags. A comparison of the whole-library genes' expression patterns suggested four groups: i) mycelia and zoosporangia, ii) zoospores and cysts, iii) germinating cysts, and iv) five infection site libraries (IF1.5 to IF24h). The libraries from the different groups showed major transitional shifts in gene expression. From the ten libraries, 722 gene expression–pattern clusters were obtained and the top 16 clusters, containing more than half of the genes, comprised enriched genes with different functions including protein localization, triphosphate metabolism, signaling process, and noncoding RNA metabolism. An evaluation of the average expression level of 30 pathogenesis-related gene families revealed that most were infection induced but with diverse expression patterns and levels. A web-based server named the *Phytophthora* Transcriptional Database has been established.

Phytophthora sojae is an oomycete plant pathogen that causes stem and root rot of soybean. The oomycetes are fungus-like organisms that are evolutionarily related to algae and are classified in the kingdom Stramenopila (Baldauf et al. 2000; Forster et al. 1990; Harper et al. 2005). The economic impact of *P. sojae* is large, as it is responsible for $1 to 2 billion in crop losses per year worldwide (Tyler 2007). Thus, this organism has been the focus of molecular genetic and genomic studies and is a model species for the study of oomycete plant pathogens (Tyler 2007), along with the potato and tomato pathogen *Phytophthora infestans* and the *Arabidopsis thaliana* pathogen *Hyaloperonospora arabidopsidis* (Coates and Beynon 2010).

P. sojae has a narrow host range and is restricted primarily to soybean (Erwin and Ribeiro 1996). It is a homothallic organism that propagates clonally and through rare sexual outcrossing (Tyler 2007). Asexual single-celled zoospores are biflagellate, motile, and chemotactic to isoflavonoids secreted by soybean roots (Morris et al. 1998; Hua et al. 2008; Tyler 2002). Zoospores encyst and germinate on the root or hypocotyl surface, and the resulting germ tube may swell to form an appressorium-like structure at the point of penetration into host tissues (Moy et al. 2004; Tyler 2007). Soybean cultivars that carry an effective resistance (*Rps*) gene to an attacking *P. sojae* strain react rapidly with a hypersensitive response (HR), which is activated within hours of zoospore attachment and arrests further pathogen growth. This is characteristic of a resistant or incompatible interaction. In contrast, no early HR occurs in a susceptible or compatible interaction, and *P. sojae*, which is hemibiotrophic, is able to colonize host cells in an initial biotrophic phase of growth that lasts for approximately 12 h (Moy et al. 2004). At later stages of infection, the pathogen enters a necrotrophic growth mode, spreading quickly throughout host tissues, causing large, water-soaked, necrotic lesions and leaving dead host cells in its wake (Supplementary Fig. S1).

Defining the *P. sojae* transcriptome is an important molecular strategy to study gene function and to dissect the molecular events that accompany pathogenesis. Gene transcripts can be profiled by high throughput techniques such as serial analysis of gene expression (Velculescu et al. 1995), microarray (Lockhart et al. 1996; Schena et al. 1995), and sequencing of clones from cDNA libraries (Adams et al. 1995; Asmann et al. 2009; Boguski et al. 1994). For the last decade, expressed sequence tag (EST) analysis and oligo-nucleotide microarrays have been relied upon for transcriptional profiling of *P. sojae* and its interaction with the soybean host plant. At least 31,314 *P. sojae* EST have been generated from a variety of tissues and conditions, including free-swimming zoospores, germinating cysts, mycelium, and *P. sojae*-infected soybean tissue (Qutob et al. 2000; Torto-Alalibo et al. 2007). An amplified-cDNA microarray containing 3,927 soybean and 969 *P. sojae* sequences was constructed to examine plant and pathogen gene expression over a timecourse of infection (Moy et al. 2004; Tyler 2007). More recently, a commercial array containing 37,637 soybean and 15,820 *P. sojae* targets was used to profile transcription during infection (Dong et al. 2009; Qutob et al. 2009; Zhou et al. 2009).

The published 95-Mbp *P. sojae* genome and its 19,027 predicted gene models (Tyler et al. 2006) provide important reference points for more comprehensive transcriptional studies that can be done using next-generation sequencing (NGS) technology (Mardis 2008). The NGS technologies offer an opportunity to exhaustively sample transcripts and digitally measure transcription levels in particular organs, tissues, or cells, under different treatments or conditions (Asmann et al. 2009). For example, the 3′-tag digital gene expression (DGE)

Corresponding author: Yuanchao Wang; Telephone: +1 86-25-84399071; Fax: +1 86-25-84395325; E-mail: wangyc@njau.edu.cn

*The ***e*-Xtra** logo stands for "electronic extra" and indicates that six supplementary figures and eight supplementary tables are published online.

protocol generates such extensive sequence data and depth-of-coverage that even the rare transcripts can be detected and quantified. This method uses oligo-dT to generate libraries that are enriched in the 3′ untranslated regions of polyadenylated mRNAs and produces 20- to 21-bp cDNA tags (Eveland et al. 2010; Morrissy et al. 2009; t Hoen et al. 2008). The expression level of virtually all genes in the sample is measured by counting the number of the tags produced from each gene (Xiang et al. 2010). Previously, the DGE profiling of transcriptomes for organisms with completed genomes confirmed that the relatively short reads produced can be effectively assembled and used for comparison of gene expression profiles (Hegedus et al. 2009; Rosenkranz et al. 2008; Wang et al. 2010).

In this study, we report the massive parallel sequencing of *P. sojae* transcriptome by the DGE protocol. A total of ten RNA samples from various life cycle stages of development and infection were sequenced, and the results were analyzed. A web server named the *Phytophthora* transcriptional database was developed to provide access to this transcriptome data.

RESULTS

Ten stages of development and infection were sampled for library sequencing.

To capture a variety of developmental and infection stages, a total of ten samples were collected from *P. sojae* P6497. The five axenically grown stages were mycelia (MY), zoosporangia (SP), zoospores (ZO), cysts (CY), and germinating cysts (GC) (Fig. 1). The five infection stages, 1.5, 3, 6, 12, and 24 h after inoculation onto susceptible soybean leaf tissues (IF1.5h to IF24h), are illustrated in Figure 1, and the description of each infection stage is provided in Supplementary Table S1. On the basis of the Illumina 3′-tag DGE protocol, we generated between 3.9 and 14.8 million raw tags for each of the ten samples. After removing low-quality reads, the total number of clean tags per library ranged from 3.7 to 14.0 million and the number of tag entities with unique nucleotide sequences (distinct tags) ranged from 84,181 to 666,004 (Table 1). From all libraries, 1,585,220 distinct tags were obtained, with a total of 90.1 million clean tags (Supplementary Fig. S3A).

Large proportions of *P. sojae* genes were detected in the ten libraries.

The *P. sojae* reference genome assembly and the 19,027 predicted genes were used as a reference sequence database (Joint Genome Institute website, *P. sojae* v1.1) (Tyler et al. 2006). The reference database contains 486,711 distinct tags. After mapping the clean tags generated from the ten DGE libraries to the reference database, at least 68.4% of the clean tags were mapped to the reference database in seven of the libraries (MY, SP, ZO, CY, GC, IF1.5h, and IF3h). The mapped percentages were lower (between 35.6 and 40.2%) in the remaining three libraries (IF6h, IF12h, and IF24h) because the samples also contained soybean tissue. Next, the gene-expression levels were determined by calculating the number of tags for each gene and, then, normalizing this value to the number of transcripts per million tags (TPM) (Asmann et al. 2009; Wang et al. 2010). A large portion of the 19,027 predicted *P. sojae* genes was detected ranging from 10,532 (55.4%) to 13,805 (72.6%) among the ten libraries (Table 1; Fig. 2A). When considering all libraries, 14,969 (78.7%) *P. sojae* genes were detected in at least one library (Fig. 2A).

Detected genes are assigned reliability categories based on a multiple loci of the genome (MMT) value.

In this study, clean tags mapped to the reference database from MMT were used to calculate a reliability measure, called the MMT value, based on the proportion of TPM derived from MMT. Because the data of MMT can not be separated to individual mapped genome loci, they only represent the collective expression levels of all genome loci having the identical tag sequence. Another type of tag, unique mapped tags (UMT), unambiguously mapped to single genome loci, indicating that the transcription data are reliable. According to this, from the ten libraries, 10,044 (52.8%) were detected by perfect matches to UMT and were marked as reliable (MMT value = 0%; unless otherwise stated, this set of genes was used for all further analyses in this research), 4,915 (25.8%) were detected but the expressed tags potentially matched MMT and were, therefore, marked as unreliable (MMT value > 0%) and 3,134 (16.5%) were not detected (Fig. 2A; Supplementary Table S2). The remaining 934 genes (4.9%) lack a tag site in their sequences to be mapped by DGE-generated tag, rendering them

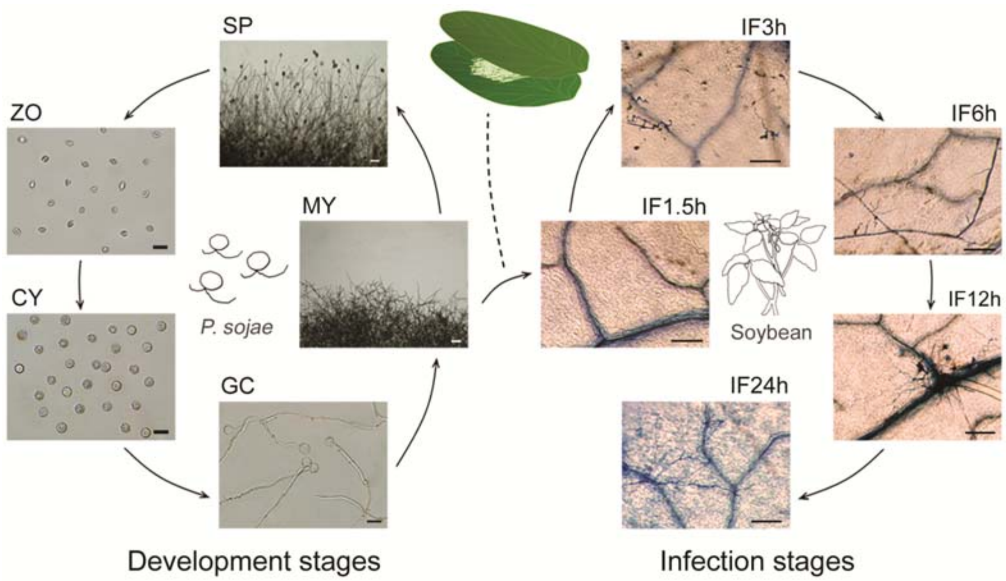

Fig. 1. Schematic illustration and microscopic observation of the ten sampled stages: mycelia (MY), zoosporangia (SP), zoospores (ZO), cysts (CY), germinating cysts (GC), and IF1.5h to IF24h (samples from 1.5, 3, 6, 12, and 24 h after infection of soybean leaves). The sandwiched inoculation method is shown in the center of the figure. The scale bars in CY, ZO, and GC are 20 µm, others are 100 µm.

undetectable by our sampling method. The full lists of the 18,093 genes having at least one tag are provided in Supplementary Tables S4 and S5.

Comparison of DGE values with quantitative reverser transcription-polymerase chain reaction (qRT-PCR) analysis results.

Total sequence tags from all libraries that match to the 10,044 genes classified as detected and reliable were plotted as integrated $\log_2$ values. As shown in the graph in Figure 2B, the highest number of genes (1,737) distributes around 7 (≥7 and <8, and 7 refers to 128-1 TPM) and 9.0% of the genes (899 of 10,044) are highly expressed at more than 10. To determine whether gene expression levels estimated by TPM counts were comparable to qRT-PCR results, 28 genes from different families and with a range of expression values were selected for further study. These genes and primers are listed in Supplementary Table S3. Expression levels of the 28 genes were determined by qRT-PCR from the ten different RNA samples used for the DGE analysis, resulting in 280 datapoints (Fig. 2C). The Pearson correlation coefficient (R value) between the cycle threshold (Ct) value of the qRT-PCR analysis and the $\log_2$ TPM values of the DGE analysis was –0.75, meaning that the genes' expression levels from DGE analysis are positively correlated with qRT-PCR (lower Ct value refers to higher expression level). This correlation between the two different platforms ($|R|$ = 0.75) is higher than many correlations of whole-library gene-expression patterns between two DGE libraries (but not replication) in our study but also lower than many others, such as IF3h to IF6h (R = 0.97). A Pearson correlation coefficient matrix of all library pairs was generated and is provided in Supplementary Figure S4, with R values ranging from 0.48 (ZO-IF3h) to 0.97 (IF3h to IF6h).

Correlation among the ten libraries establishes four different expression groups.

To study the relatedness of overall gene-expression patterns among the ten libraries, a hierarchical clustering (HCL) tree using the Pearson correlation method with average linkage was constructed, using the DGE data from the 10,044 detected reliable genes (Fig. 3A). This shows that the DGE profiling of the ZO library is closest to CY, GC is alone in a branch, MY is close to SP, and the five infection site libraries (IF1.5h to IF24h) exist in the same branch. Principal component analysis

Table 1. Summary of the output data and mapping work

Tag or gene name[a] Category[b]	MY[c]	SP	ZO	CY	GC	IF1.5h	IF3h	IF6h	IF12h	IF24h
Raw tags										
Total	7,618,146	3,889,220	9,115,349	8,722,222	10,942,428	14,794,423	10,406,821	11,338,625	10,353,209	10,660,949
Distinct	339,251	235,351	305,957	341,113	1,400,829	980,071	1,439,380	2,052,365	1,917,624	2,011,485
Clean tags										
Total	7,406,626	3,737,336	8,909,709	8,491,318	9,936,835	14,033,907	9,387,619	9,930,713	9,010,412	9,247,642
% of raw tags	97.2%	96.1%	97.7%	97.4%	90.8%	94.9%	90.2%	87.6%	87.0%	86.7%
Distinct	128,354	84,181	102,771	112,284	432,812	286,150	451,475	666,004	596,675	616,298
% of raw tags	37.8%	35.8%	33.6%	32.9%	30.9%	29.2%	31.4%	32.5%	31.1%	30.6%
Clean tags mapping to genome or gene										
Total	6,763,028	3,397,523	8,221,308	7,861,599	6,933,838	12,980,742	6,421,067	3,987,978	3,514,638	3,287,787
% of clean tags	91.3%	90.9%	92.3%	92.6%	69.8%	92.5%	68.4%	40.2%	39.0%	35.6%
Distinct	95,904	65,520	74,605	79,620	186,826	212,225	191,088	155,151	145,068	141,085
% of clean tags	74.7%	77.8%	72.6%	70.9%	43.2%	74.2%	42.3%	23.3%	24.3%	22.9%
All tag-mapped genes										
gene	12,323	11,978	10,532	11,450	13,179	13,805	12,743	12,519	12,394	12,196
% of 19,027	64.8%	63.0%	55.4%	60.2%	69.3%	72.6%	67.0%	65.8%	65.1%	64.1%

[a] Raw tags, sequence data prior to trimming and processing; clean tags, trimmed and processed 21-bp sequences.
[b] Distinct tags are classified according to their sequence. The *Phytophthora sojae* reference genome (Joint Genome Institute *P. sojae* v1.1) describes 19,027 predicted genes, including 18,093 with at least one tag.
[c] MY = mycelia, SP = zoosporangia, ZO = zoospores; CY = cysts, GC = germinated cysts, and IF1.5h to IF24h, indicates samples from 1.5, 3, 6, 12, and 24 h after infection of soybean leaves.

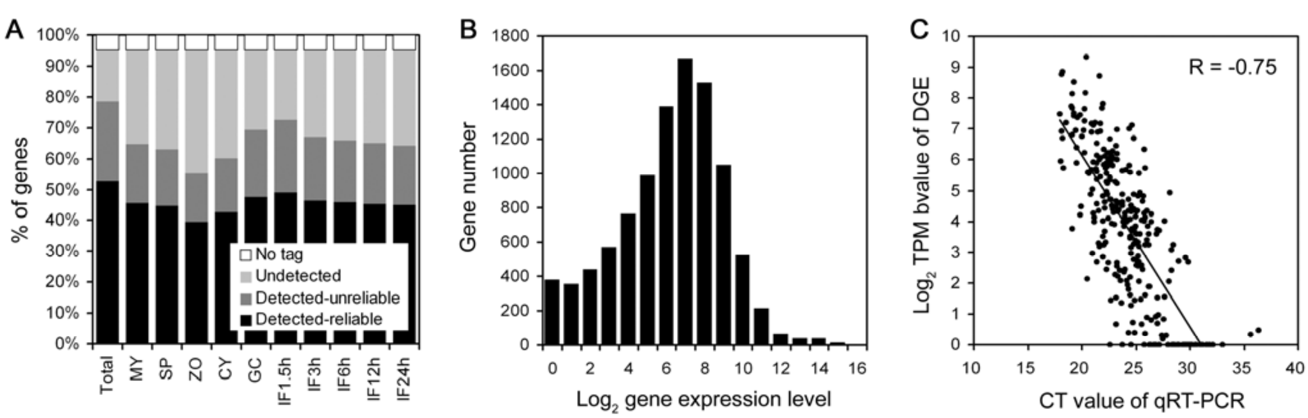

Fig. 2. Characteristics and validation of detected genes. **A,** The distribution of genes within the different detectable categories, based on the sequence tag analysis as described in the text. **B,** The distribution of gene expression levels, based on the number of genes falling in each $\log_2$ gene expression category. Data are from all ten samples. **C,** Validation of digital gene expression (DGE) data by quantitative reverse transcription-polymerase chain reaction (qRT-PCR). Scatter plots indicate the cycle threshold (Ct) value of qRT-PCR analysis and the $\log_2$ transcripts per million tags value of DGE for 280 datapoints from 28 genes in ten samples. The Pearson correlation coefficient (R) is also shown.

(PCA) is a statistical method to reduce the dimensionality of the dataset and allows a visual inspection of the samples based on gene-expression profiles. Samples with a similar gene-expression profile would cluster in the same direction (Elferink et al. 2011). The PCA plot of principal component (PC) 1 and PC 2 shown in Figure 3B reveals a situation consistent with the HCL tree. In the PCA plot, the libraries in the four major branches of the tree (ZO and CY, GC, MY and SP, and IF1.5h to IF24h) locate in distinguishably different regions. The accumulated eigenvalue of the first two PC is 74.2% (PC 1, 56.7% and PC 2, 17.5%), which means that the information from PC 1 and PC 2 have a highly reliable degree.

Differentially expressed gene analysis reveals transcriptional shifts between libraries.

To study the differentially expressed genes between each library pair, we performed filtering to identify twofold upregulated and twofold downregulated genes with P value ≤ 0.01, employing the $Chi^{2'}$ test and Bonferroni correction. The results, shown in Figure 4A, indicate that the greatest changes in gene expression occur during cyst germination and host infection, when thousands of genes were detected as upregulated (GC and IF1.5h to IF24h compared with MY, SP, ZO, and CY). However, by contrast to the five infection libraries, GC downregulated more genes compared with MY and SP and upregulated fewer genes compared with ZO and CY. This also confirmed that the GC was a distinct group. For ZO and CY, many more genes were downregulated when compared with the other libraries. In contrast, comparison among the infection libraries (IF1.5 to IF24h) indicates that gene-expression patterns changed steadily but not dramatically during the course of infection. The full list of differentially expressed genes between each library pair is provided in Supplementary Table S6. Three stage pairs (MY-ZO and ZO-IF1.5h, with the largest number of genes down- or upregulated, and IF3h-IF6h, with least genes changed) were selected to represent the distribution of genes corresponding to different fold change categories (Fig. 4B).

Clustering analysis shows different gene-expression patterns.

The ten libraries provide a wide range of stages to understand gene expression during the *P. sojae* life cycle. To elucidate detailed gene-expression patterns, the clustering affinity search technique (CAST) was used to generate clusters (Saeed et al. 2003). Figure 5A shows a breakdown of 722 clusters generated with members (genes) ranging from 1,675 to 1. Most of the clusters (662 clusters) had no more than 20 genes. However, the top 16 clusters, each of which had more than 90 members, contained >50% (51.4%) of all detected genes, illustrating the major gene-expression patterns (Fig. 5B). Among the top 16 clusters, different sets of genes were upregulated at varied stages: clusters a, b, and i, during infection; k and g, early infection; o, middle infection; n and p, late infection; e, during development; d, both MY and SP; c and m, both ZO and CY; j, ZO, CY and GC; and f, h, and l were stage-specific at GC, SP, and IF1.5h, respectively. The clusters with similar patterns mentioned above also showed differences in detailed stages. For example, the infection-related genes in cluster b were up-regulated in stages from GC to IF24h but, in cluster i, were from the later stages (IF1.5h to IF24h).

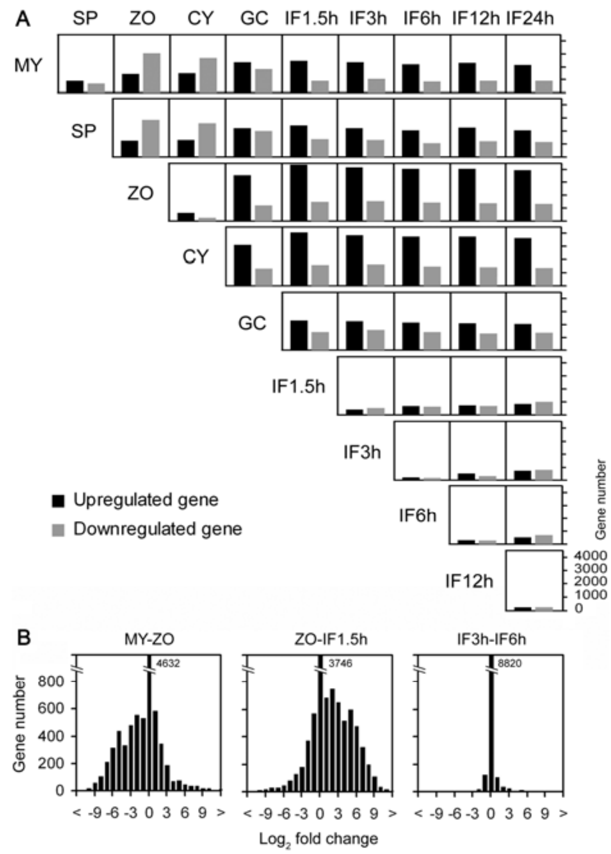

Fig. 4. Differentially expressed genes of each of two libraries. **A,** The number of upregulated and downregulated genes in each of two libraries. Differentially expressed genes are identified by filtering of the twofold up- and downregulated genes with $P \leq 0.01$, employing both the $Chi^{2'}$ test and Bonferroni correction. **B,** The distribution of $\log_2$ fold change levels for three selected stage pair–wise (MY-ZO and ZO-IF1.5h, with the largest number of genes down- and upregulated; IF3h-IF6h, the most stable stage pair).

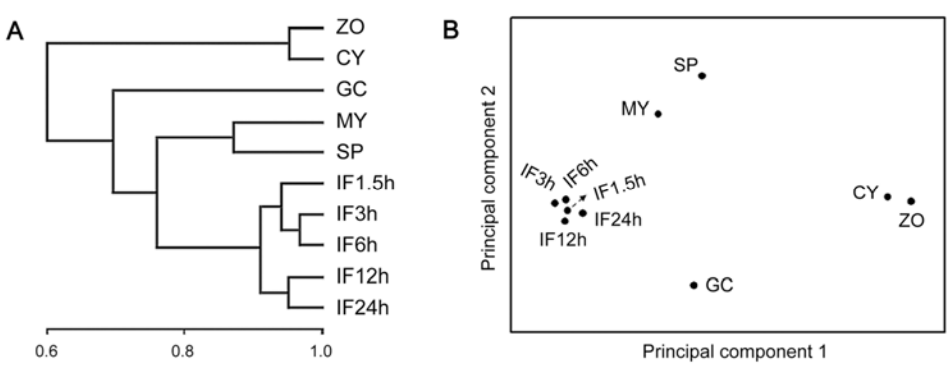

Fig. 3. Correlation of the whole-library genes expression patterns. **A,** Hierarchical clustering tree. The node height scale is shown below. **B,** Principal component analysis plot of principal components 1 and 2, whose eigenvalues are 56.7 and 17.5%, respectively. The analyses for A and B were performed by the MultiExperiment Viewer (vs. 4.6) software, using the 10,044 genes expression data.

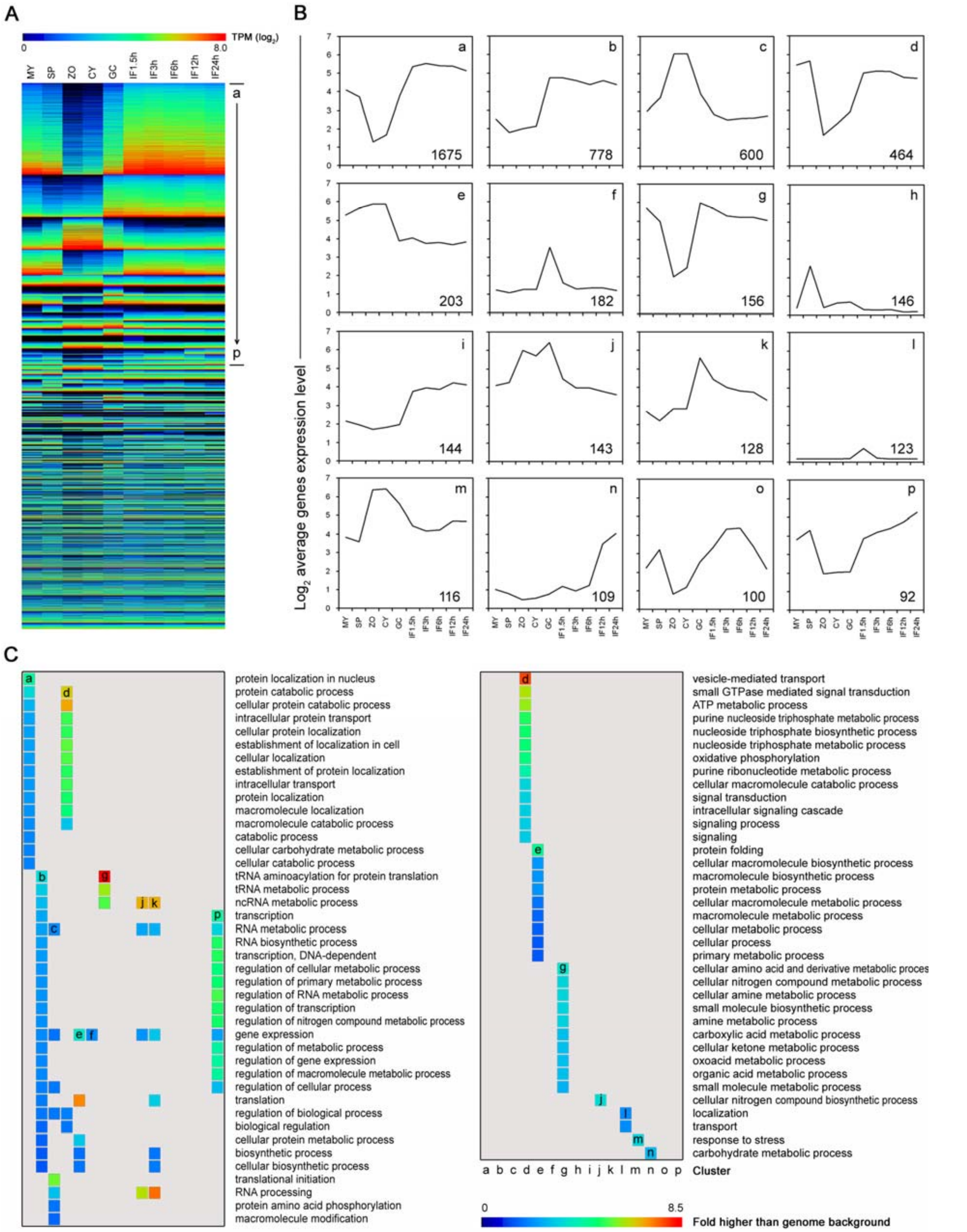

Fig. 5. Clustering and gene ontology (GO) enrichment analysis of gene expression patterns. **A,** Heat map shows the 722 gene–expression clusters generated by the clustering affinity search technique method. Each line refers to data of one gene. The order is from the cluster with the most members (1,675 genes) to that with the least members (a single gene). The color bar represents the $\log_2$ of transcripts per million tags values, ranging from dark blue (0) to red (8.0). **B,** Log_2 average gene-expression levels of the top 16 clusters. The number of cluster members is marked at the bottom right of each plot. **C,** GO enrichment analysis of genes from the top 16 clusters. The mapped GO terms referring to biological processes were compared with the whole genome (GO terms for all of the 19,027 genes) background and were filtered with P value ≤ 0.5 by $\mathrm{Chi}^{2'}$ test and false discovery rate correction. The color bar represents the fold higher than genome background of GO terms proportion, ranging from dark blue (0) to red (8.5). The gray blocks mean no data or data that were filtered by the above criteria. The cluster names were marked for the first block at each column for distinguishing.

Functional annotation of genes in clusters with similar expression patterns.

To study the major gene functions of different expression patterns, we performed gene ontology (GO) enrichment analysis for genes from the top 16 clusters (Fig. 5B). The mapped GO terms referring to biological process were compared with the whole-genome background (GO terms for all of the 19,027 predicted *P. sojae* genes) and were filtered with P value ≤ 0.5 by Chi$^{2'}$ test and false discovery rate (FDR) correction. Figure 5C shows that the number of enriched GO terms per cluster ranged from 0 (clusters h, i, and o) to 26 (cluster d). Different gene functions were overrepresented in certain clusters. For example, the genes in cluster a were rich in function of protein localization (transport). Cluster d also had this characteristic, but additionally, it had genes related to processes such as the triphosphate metabolic process and the signaling process. Genes in clusters b and p mostly referred to regulation of metabolic process and transcription. A GO term, noncoding RNA metabolic process, was found in clusters b, g, j, and k, whose expression patterns were all infection related. Furthermore, for clusters m and n, a single enriched GO term was matched to response to stress and carbohydrate metabolic process, respectively. The full list of enriched GO terms and corresponding genes was provided in Supplementary Table S7.

Average gene-expression patterns of putative pathogenicity gene families.

Many gene families from *P. sojae* were studied or suggested to have important roles in pathogenesis. To further understand the expression patterns, 30 different gene families or groups were selected, with each group containing from three to 396 distinct genes (Torto-Alalibo et al. 2007) (Fig. 6). To determine an average expression pattern, the TPM values for each stage were calculated group by group by pooling data from the UMT and the nonredundant MMT. The MMT provide good data for this purpose because this is an analysis of gene-family expression, anticipating that many MMT were mapped to the same group. The results illustrated diverse expression patterns but most gene groups were up-regulated during the above-mentioned major transcriptional shifts. For example, the PDR-like ABC transporters, glutathione transferase, and glutaredoxin were up-regulated at GC; the aspartyl proteinases were highly expressed at GC and IF1.5h; the cutinases were up-regulated from ZO and highly expressed in GC; the RxLR and NLP effector families were highly induced at GC and with another peak at late infection; another two effector families, CRN and elicitin (or elicitin-like), although expressing similar two-peak patterns, consistently showed higher average expression levels than RxLR and NLP; however, elicitin was expressed with higher level at MY, ZO, and CY. For ubiquitin protease, the genes stably had a high expression level.

Community access to the *Phytophthora* transcriptional database (PTD).

We established the PTD web server (v1.1) to allow the research community easy access to our data (Supplementary Fig. S5). Each gene has a detailed page describing its basic annotation and the DGE transcriptional data. The transcriptional data

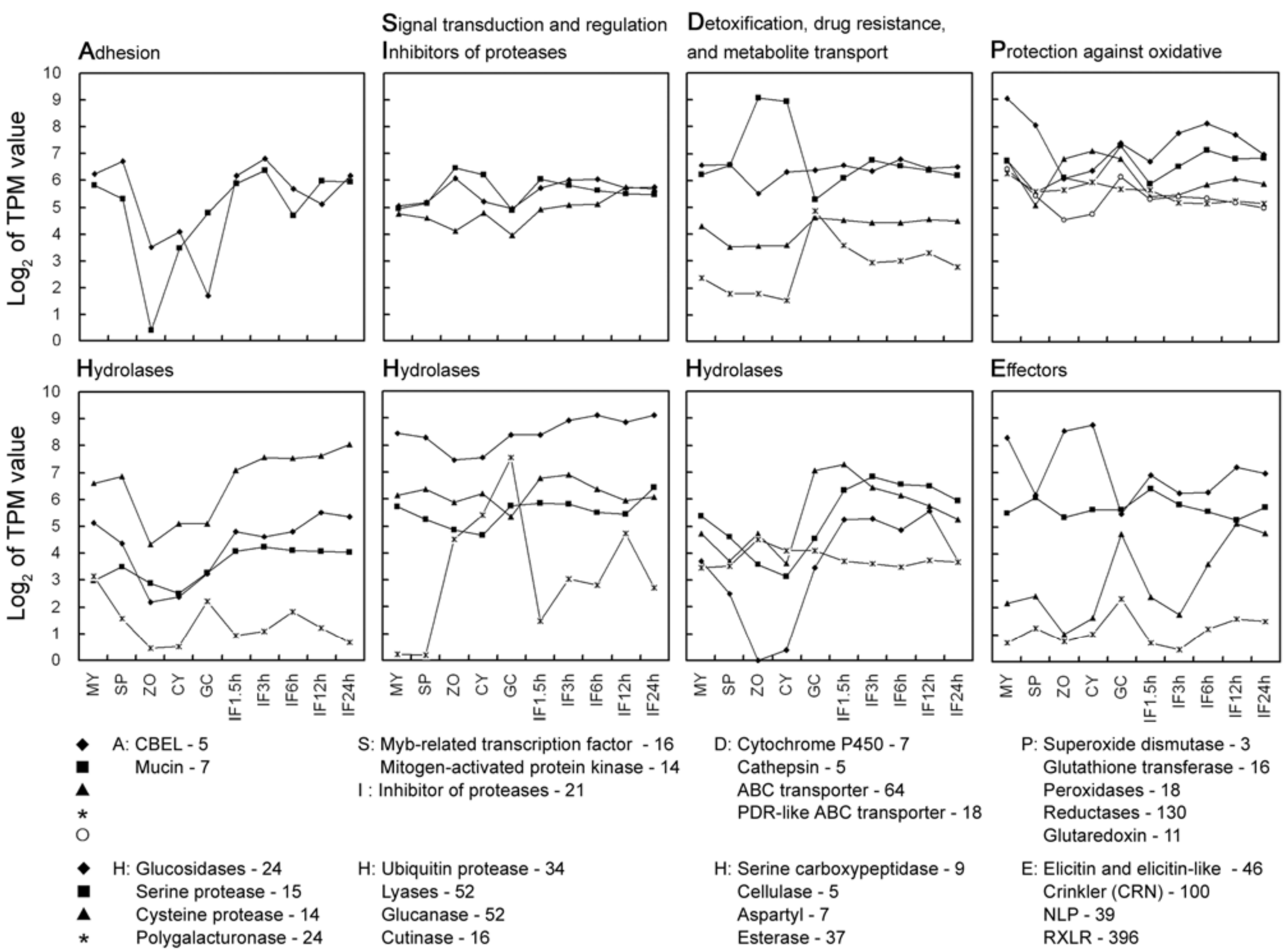

Fig. 6. Average gene expression levels of the 30 putative pathogenicity gene families. The legend shown on the bottom of the figure identifies the gene families plotted in each graph. The number of genes included in calculating average expression values is shown at the right side of family name. The function categories are marked at the top of each figure.

include expression TPM values together with MMT values and sequences that correspond to the gene of interest. Graphical outputs include histogram views that are intuitive for exploring the data. The PTD provides access to the gene data via searches by the gene ID, annotation, fold change between two stages, and BLAST to search sequence-homologous genes. It also provides links to the *P. sojae* databases and assemblies at the Department of Energy Joint Genome Institute (Tyler et al. 2006) and the Virginia Bioinformatics Institute microbial database (Tripathy et al. 2006) for quick access to extensive gene or genome contextual information. To facilitate further analysis of these data, the related tag data and analysis results are also provided for downloading. A list of the current web resources related to oomycete genomics research is also provided in PTD and Supplementary Table S8.

DISCUSSION

In this study, we used massively parallel sequencing technologies coupled with computational DGE analysis to characterize the transcriptome of *P. sojae* during development and infection. Our results provide an extensive picture of transcription in *P. sojae* and offer investigators a rich set of sequence data for reference and interrogation.

To organize the voluminous data and to differentiate the transcripts, we based our computational analysis on 21-bp tags because this approach has been proven successful in other species (Asmann et al. 2009; Saha et al. 2002). Nevertheless, the numerous gene paralogs sharing sequence tags resulted in a high frequency of MMT. The detected genes with MMT represented 25.9% of the total number of genes detected in *P. sojae*. In some DGE studies MMT are filtered during the mapping process (Asmann et al. 2009; Wang et al. 2010). Such an approach provides expression values for all reliably tagged genes. However, the expression values for genes with MMT may be underestimated or completely lacking; these genes would correspond to genes detected-reliable or undetected in our study (2,324/12.2% and 2,591/13.6% genes fit these descriptions, respectively). To avoid this problem, the MMT data were counted for gene expression level, and the proportion of value derived from MMT was used to calculate an MMT-value as a reliability measure. This method provides more complete information of gene expression level and allows for the classification of reference genes into four categories: detected-reliable, detected-unreliable, undetected, and no tag. Although we deemed genes with MMT to be detected-unreliable and did not analyze these genes in this study, their expression data and MMT-values were also provided in PTD, which are still valuable for genes with low MMT-values. Moreover, the MMT data remain useful for determining collective expression patterns for the different genes sharing the same tags (Fig. 6).

Another limitation of DGE analysis lies in the annotation of the reference genome. The 19,027 predicted *P. sojae* genes represent an estimation based on gene modeling programs (Tripathy et al. 2006; Tyler et al. 2006). Many gene models are erroneous, and annotation may be completely lacking for certain genes. These would cause the bias or absence of gene expression data. However, the transcript tag data that we have developed may be used to improve gene models (Supplementary Fig. S6). It is even possible that tags that do not map to the genome or to a predicted gene represent the junction of two exons in a misannotated gene. Genome annotation is an iterative process that will inevitably improve with time. The sequence data generated in this study offer an opportunity to improve the annotation of the *P. sojae* genome. Biologists have long realized that problems with gene annotation exist for even the best-characterized model species; thus the situation for *P. sojae* is not unusual or extraordinary.

Besides the above mentioned limitations, there are still several problems with DGE, including statistical modeling (particularly normalization) (Mak 2011); the depth of sequencing required to effectively sample the transcriptome; the cost, which may tempt some to avoid using biological replicates; and the bioinformatics required to manage such a large amount of data (Malone and Oliver 2011). However, every technology has its advantages and inherent biases. For example, of the established methods, microarrays remain useful and accurate tools for measuring expression levels (Malone and Oliver 2011) with high throughput, but they have relative low sensitivity for the detection of rare transcripts and potentially can miss many targets that may not be included on the array. EST studies can obtain longer full-length transcripts; however, they are incomplete in their coverage of transcripts and are expensive to perform (Asmann et al. 2009). Although the DGE transcriptome profiling also has some draw-backs, it is based on the produced relatively short reads and used for comparison of gene expression profiles (Hegedus et al. 2009; Rosenkranz et al. 2008; Wang et al. 2010), with extensive sequence data and depth-of-coverage such that even the rare transcripts can be detected and quantified (Eveland et al. 2010; Morrissy et al. 2009; t Hoen et al. 2008). Thus, NGS-based transcription profiling methods can complement and extend other technologies, (Malone and Oliver 2011), but it will take time to update and improve the technology and protocol.

To evaluate the DGE data, a comparison with 280 datapoints from the output of two different platforms (DGE and qRT-PCR) was performed that indicated a positive correlation between the data. And the correlations between DGE data from similar samples indicated that the highest R value was 0.97 (IF3h-IF6h). Overlooking the transcriptional analysis of ten stages of the *P. sojae* life cycle, four library groups were found by a comparison with whole-library gene-expression patterns. This was confirmed by the analysis of differentially expressed genes, which showed major transitional shifts between the libraries from the different groups. Beside the inherent technological problems of DGE, some other points cannot be excluded when interpreting the data, including the experiment time, conditions, and people. However, these transcriptional shifts generally agreed with the biological process, for between the libraries, that of the pathogen grown axenically (MY, SP, ZO, CY, and GC) and that of the pathogen grown in contact with the plant (IF1.5h to IF24h), the host-pathogen interaction can reasonably be expected to modify the pathogen's expression profile. The distinction of ZO and CY from the other libraries is probably due to its divergence from non-mycelium status (e. g., they are either swimming in water or attached to the host surface). And GC is a transitionary status between cysts and mycelium. For the RNA samples we used comparing the seven libraries collected from the pathogen only, we also paid attention to the other three timepoint samples (IF6h, IF12h, and IF24h), collected from a mixture of inoculated mycelium and host tissue. In these three libraries, fewer clean tags were mapped to the reference sequences, although there was no obvious difference in the number of detected genes (Table 1; Fig. 2A) or in the distribution of gene numbers assigned to different expression level (data not shown). The whole-library expression patterns and the analysis of differentially expressed genes for these three samples also did not show distinction from the unmixed samples IF1.5h and IF3h (Figs. 3A and B and 4A).

P. sojae is a widespread and destructive plant pathogen and is one of the best-studied species among oomycete organisms. Here, we have demonstrated that deep sequencing of the transcriptome combined with computational tools such as DGE can provide an unparalleled level of detail and coverage of gene-expression patterns. Based on these DGE data and other available resources of *P. sojae*, a number of questions remain for further study. For example, what are the identities of the genes and how do their functions contribute to the transcriptional shifts or specific expression patterns at certain stages? Is there really a set of noncoding RNAs with important roles in the interaction of *P. sojae* with its host? Are there novel gene families playing important roles in the life of *P. sojae* that could be found according to the species-specially expansion of gene copy numbers and the expression patterns, e.g., the cutinase family? Finally, the pathogen effector is one of the hot spots in current research of pathogen-host interaction. Further studies on the effector genes, including RxLR, CRN, NLP, and even unknown genes are on the way. In brief, these data will serve as a valuable public genomic resource and will help further clarify the biology of *Phytophthora* plant pathogens.

MATERIALS AND METHODS

Preparation of biological material.
P. sojae P6497 (race 2) (Forster et al. 1994), from which the reference genome was derived (Tyler et al. 2006), was used in this research. MY were cultivated in 10% V8 liquid medium at 25°C in darkness for 48 h and were then blotted dry with absorbent paper and were preserved in liquid nitrogen for RNA isolation. SP were induced by repeatedly washing 48- to 72-h-old mycelial mats with sterile distilled water at 25°C in darkness until sporangia formed abundantly. ZO were released by placing the zoosporangial mycelial mat into 10 ml of sterile distilled water at 5 to 10°C for 10 to 15 min, and then, at 25°C for 10 to 30 min. The ZO were then concentrated by centrifugation at 2,000 rpm at 0°C to a concentration of >150 ZO per microliter and were then preserved in liquid nitrogen for RNA isolation. The ZO were counted under a microscope based on 2 µl of concentrated ZO suspension sample with three repeats. CY were produced by vortexing the ZO suspension at room temperature for 30 s and were then collected by centrifugation at 2,000 rpm at 0°C and were preserved in liquid nitrogen for RNA isolation. GC were obtained by cultivating cysts with 5% V8 liquid medium at 25°C, 150 rpm for 1 h and were then collected by centrifugation at 2,000 rpm at 0°C. For mycelial infection (IF1.5h to IF24h), the soybean cultivar Williams, which is susceptible to P6497, was grown in a greenhouse at 22 to 28°C and was used at the second-leaf stage. Soybean leaves were treated with 0.05% vol/vol solution of Tween 20 to improve wetting. A mycelial mat was washed with sterile distilled water and was then laid on and sandwiched (Fig. 1) between upper surfaces of two leaves at 25°C, respectively, for 1.5, 3, 6, 12, and 24 h after infection. For the 1.5- and 3- h timepoints, the mycelial mat was carefully peeled from the leaves and preserved in liquid nitrogen. For the later timepoints, the regions of the leaves in contact with the mycelia were excised together with the mycelia and were preserved in liquid nitrogen. Parallel samples were prepared simultaneously and were used for microscopic analysis. In addition to the infection samples, the leaves were decolorized with absolute ethanol before observation.

Library preparation and sequencing.
Tag library preparation for the ten samples was performed in parallel, using the Illumina gene expression sample preparation kit. Each sample of 6 µg of the total RNA was extracted from above-mentioned samples (Total RNA purification system; Invitrogen, Carlsbad, CA, U.S.A.), and mRNA was purified by oligo (dT) magnetic bead adsorption. mRNA bound to the beads was then used as a template for first-strand cDNA synthesis primed by oligo (dT), and the second-strand cDNA was consequently synthesized using random primers. The cDNA was cleaved with *Nla*III at CATG sites, and then, the cDNA fragments with 3′ ends were purified with magnetic beads and Illumina adapter A was added to their 5′ ends, creating a recognition site of *Mme*I at the junction. *Mme*I cleaves 17 bp downstream of the CATG site, producing tag fragments that include Illumina adapter A. After removing 3′ fragments by magnetic bead precipitation, Illumina adapter B was introduced at the 3′ ends of tags, thus acquiring tags with different adapters at each end to form a tag library. After 15 cycles of linear PCR amplification, 85-bp oligonucleotides were purified by 6% Tris-borate-EDTA polyacrylamide gel electrophoresis. These oligonucleotides were then digested, and the single-chain molecules were fixed onto the flow cell (Illumina Sequencing Chip) for sequencing. Raw reads were generated with a sequencing length of 35 bp (Supplementary Fig. S2).

Analysis and mapping of DGE tags.
Raw sequences have 3′ adaptor fragments as well as a few low-quality sequences and several types of impurities. Raw sequences were transformed into clean 21-bp (CATG+17 bp) tags by the following steps: i) 3′ adaptor sequence was trimmed, resulting in 21-bp tags from 35 bp of raw sequence, ii) empty reads were removed (reads with only 3′ adaptor sequences but no tags), iii) low-quality tags were removed (tags with ambiguous base calls), iv) tags of unusual length were removed, leaving only tags of 21 bp, and v) nonredundant tags were removed (each tag needs to be detected at least twice to be considered reliable). These raw datasets are available at the National Center for Biotechnology Information Gene Expression Omnibus database with the accession number GSE29651. A preprocessed database of all possible CATG+17 bp tag sequences was created using the *P. sojae* genome and gene models as a reference database (*P. sojae* v1.1) (Tyler et al. 2006). All clean tags were mapped to this reference database, allowing no more than 1 bp mismatch. The number of mapped clean tags were calculated for each library and were then normalized to TPM. Clean tags mapped to reference sequences from multiple loci (MMT) were identified but not removed. The expression value for a gene was derived from the sum of TPM for all mapped tags. The proportion of TPM derived from MMT was used to calculate an MMT value.

SYBR green real-time RT-PCR assay.
A total of 28 genes were selected for SYBR green real-time RT-PCR assay, each using the same RNA for DGE from ten samples, resulting in 280 datapoints. Pearson correlation coefficient was calculated between the Ct value of the qRT-PCR analysis and the $\log_2$ TPM values from the DGE analysis. A real-time RT-PCR reaction (20 µl) included 20 ng of DNA, 0.2 µM each prime, 10 µl of SYBR Premix Ex*Taq* (TaKaRa Bio Inc. Shiga, Japan), and 6.8 µl of distilled H_2O. Reactions were performed on an ABI PRISM 7300 fast real-time PCR system (Applied Biosystems, Foster City, CA, U.S.A.) under the following conditions: 95°C for 30 s, 40 cycles of 95°C for 5 s, 60°C for 31 s, to calculate Ct values, followed by 95°C for 15 s, 60°C for 1 min, and then, 95°C for 15 s, to obtain melt curves. The 7300 system sequence detection software (v. 1.4) was used for data analysis.

Further analysis of gene-expression data.
The MultiExperiment Viewer (v. 4.6) software package was used to draw the heat maps, construct the HCL tree (using the Pearson correlation method with average linkage), perform the PCA (using the 'mean' for centering mode), and obtain the

gene-expression pattern clusters using CAST (the distance metric was the default Pearson correlation and the threshold affinity value was 0.9) (Saeed et al. 2003). The differentially expressed genes between each library pair were filtered by two criteria: i) two-fold over- or underrepresentation of gene expression level, and ii) P value ≤ 0.01, employing the $Chi^{2'}$ test and Bonferroni correction in the IDEG6 web server (Romualdi et al. 2003). To determine the fold change, e.g., MY-SP is calculated by $(TPM_{SP} + 0.1)/(TPM_{MY} + 0.1)$. The annotated GO terms were downloaded from *P. sojae* v1.1 at the Joint Genome Institute database. For GO enrichment analysis, the mapped GO terms referring to biological process were compared with the whole-genome background (GO terms for all of the 19,027 predicted *P. sojae* genes) and were filtered with P value ≤ 0.5 by $Chi^{2'}$ test and FDR correction using the singular enrichment analysis methods found at the AgriGO web server (Du et al. 2010).

ACKNOWLEDGMENTS

We thank B. Tyler for editing of the manuscript. This work was supported, in part, by grants to Y. Wang from NSFC (number 30671345), by the Special Fund for Agro-scientific Research in the Public Interest (3-20), and National Soybean Industrial Technology system from China; and the Agriculture and Agri-Food Canada Crop Genomics program to M. Gijzen. Y. Wang conceived the research. Y. Wang, D. Dou M. Gijzen, S. Dong, and W. Ye designed the research. W. Ye and D. Dou analyzed the data, X. Wang prepared the RNA samples, K. Tao performed the microscopic observation, Y. Lu established the PTD database, and T. Dai provided the qRT-PCR data. W. Ye, Y. Wang, and M. Gijzen wrote the paper.

LITERATURE CITED

Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Bult, C. J., Lee, N. H., Kirkness, E. F., Weinstock, K. G., Gocayne, J. D., White, O., Sutton, G., Blake, J. A., Brandon, R. C., Chiu, M., Clayton, R. A., Cline, R. T., Cotton, M. D., Hughes, J. E., Fine, L. D., Fitzgerald, L. M., FitzHugh, W. M., Fritchman, J. L., Geoghagen, N. S. M., Glodek, A., Gnehm, C. L., Hanna, M. C., Hedblom, E., Hinkle-Jr. , P. S., Kelley, J. M., Klimek, K. M., Kelley, J. C., Liu, L., Marmaros, S. M., Merrick, J. M., Moreno-Palanques, R. F., McDonald, L. A., Nguyen, D. T., Pellegrino, S. M., Phillips, C. A., Ryder, S. E., Scott, J. L., Saudek, D. M., Shirley, R., Small, K. V., Spriggs, T. A., Utterback, T. R., Weidman, J. F., Li, Y., Barthlow, R., Bednarik, D. P., Cao, L., Cepeda, M. A., Coleman, T. A., Collins, E., Dimke, D., Feng, P., Ferrie, A., Fischer, C., Hastings, G. A., He, W., Hu, J., Huddleston, K. A., Greene, J. M., Gruber, J., Hudson, P., Kim, A., Kozak, D. L., Kunsch, C., Ji, H., Li, H., Meissner, P. S., Olsen, H., Raymond, L., Wei, Y., Wing, J., Xu, C., Yu, G., Ruben, S. M., Dillon, P. J., Fannon, M. R., Rosen, C. A., Haseltine, W. A., Fields, C., Fraser, C. M., and Venter, J. C. 1995. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 377:3-174.

Asmann, Y. W., Klee, E. W., Thompson, E. A., Perez, E. A., Middha, S., Oberg, A. L., Therneau, T. M., Smith, D. I., Poland, G. A., Wieben, E. D., and Kocher, J. P. 2009. 3′ tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. BMC Genomics 10:531.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I., and Doolittle, W. F. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290:972-977.

Boguski, M. S., Tolstoshev, C. M., and Bassett, D. E., Jr. 1994. Gene discovery in dbEST. Science 265:1993-1994.

Coates, M. E., and Beynon, J. L. 2010. *Hyaloperonospora arabidopsidis* as a pathogen model. Annu. Rev. Phytopathol. 48:329-345.

Dong, S., Qutob, D., Tedman-Jones, J., Kuflu, K., Wang, Y., Tyler, B. M., and Gijzen, M. 2009. The *Phytophthora sojae* avirulence locus *Avr3c* encodes a multi-copy RXLR effector with sequence polymorphisms among pathogen strains. PLoS One 4:e5556. Published online.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. 2010. agriGO: A GO analysis toolkit for the agricultural community. Nucleic Acids Res. 38:W64-70.

Elferink, M. G., Olinga, P., van Leeuwen, E. M., Bauerschmidt, S., Polman, J., Schoonen, W. G., Heisterkamp, S. H., and Groothuis, G. M. 2011. Gene expression analysis of precision-cut human liver slices indicates stable expression of ADME-Tox related genes. Toxicol. Appl. Pharmacol. 253:57-69.

Erwin, D. C., and Ribeiro, O. K. 1996. *Phytophthora* Diseases Worldwide. The American Phytopathological Society, St. Paul, MN, U.S.A.

Eveland, A. L., Satoh-Nagasawa, N., Goldshmidt, A., Meyer, S., Beatty, M., Sakai, H., Ware, D., and Jackson, D. 2010. Digital gene expression signatures for maize development. Plant Physiol. 154:1024-1039.

Forster, H., Tyler, B. M., and Coffey, M. D. 1994. *Phytophthora sojae* races have arisen by clonal evolution and by rare outcrosses. . Mol. Plant-Microbe Interact. 7:780-791.

Forster, H., Coffey, M. D., Elwood, H., and Sogin, M. L. 1990. Sequence-analysis of the small subunit ribosomal-RNAs of 3 zoosporic fungi and implications for fungal evolution. Mycologia 82:306-312.

Harper, J. T., Waanders, E., and Keeling, P. J. 2005. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. Int. J. Syst. Evol. Microbiol. 55:487-496.

Hegedus, Z., Zakrzewska, A., Agoston, V. C., Ordas, A., Racz, P., Mink, M., Spaink, H. P., and Meijer, A. H. 2009. Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. Mol. Immunol. 46:2918-2930.

Hua, C., Wang, Y., Zheng, X., Dou, D., Zhang, Z., and Govers, F. 2008. A *Phytophthora sojae* G-protein alpha subunit is involved in chemotaxis to soybean isoflavones. Eukaryot. Cell 7:2133-2140.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. 14:1675-1680.

Mak, H. C. 2011. John Storey. Nat. Biotech. 29:331-333.

Malone, J. H., and Oliver, B. 2011. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol. 9:34.

Mardis, E. R. 2008. The impact of next-generation sequencing technology on genetics. Trends Genet. 24:133-141.

Morris, P. F., Bone, E., and Tyler, B. M. 1998. Chemotropic and contact responses of *Phytophthora sojae* hyphae to soybean isoflavonoids and artificial substrates. Plant Physiol. 117:1171-1178.

Morrissy, A. S., Morin, R. D., Delaney, A., Zeng, T., McDonald, H., Jones, S., Zhao, Y., Hirst, M., and Marra, M. A. 2009. Next-generation tag sequencing for cancer gene expression profiling. Genome Res. 19:1825-1835.

Moy, P., Qutob, D., Chapman, B. P., Atkinson, I., and Gijzen, M. 2004. Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. Mol. Plant-Microbe Interact. 17:1051-1062.

Qutob, D., Hraber, P. T., Sobral, B. W., and Gijzen, M. 2000. Comparative analysis of expressed sequences in *Phytophthora sojae*. Plant Physiol. 123:243-254.

Qutob, D., Tedman-Jones, J., Dong, S., Kuflu, K., Pham, H., Wang, Y., Dou, D., Kale, S. D., Arredondo, F. D., Tyler, B. M., and Gijzen, M. 2009. Copy number variation and transcriptional polymorphisms of *Phytophthora sojae* RXLR effector genes *Avr1a* and *Avr3a*. PLoS One 4:e5066. Published online.

Romualdi, C., Bortoluzzi, S., D'Alessi, F., and Danieli, G. A. 2003. IDEG6: A web tool for detection of differentially expressed genes in multiple tag sampling experiments. Physiol Genomics 12:159-162.

Rosenkranz, R., Borodina, T., Lehrach, H., and Himmelbauer, H. 2008. Characterizing the mouse ES cell transcriptome with Illumina sequencing. Genomics 92:187-194.

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., and Quackenbush, J. 2003. TM4: A free, open-source system for microarray data management and analysis. Biotechniques 34:374-378.

Saha, S., Sparks, A. B., Rago, C., Akmaev, V., Wang, C. J., Vogelstein, B., Kinzler, K. W., and Velculescu, V. E. 2002. Using the transcriptome to annotate the genome. Nat Biotechnol 20:508-512.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. 1995. Quantitative monitoring of gene-expression patterns with a complementary DNA microarray. Science 270:467-470.

t Hoen, P. A., Ariyurek, Y., Thygesen, H. H., Vreugdenhil, E., Vossen, R. H., de Menezes, R. X., Boer, J. M., van Ommen, G. J., and den Dunnen, J. T. 2008. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res 36:e141.

Torto-Alalibo, T. A., Tripathy, S., Smith, B. M., Arredondo, F. D., Zhou, L., Li, H., Chibucos, M. C., Qutob, D., Gijzen, M., Mao, C., Sobral, B. W., Waugh, M. E., Mitchell, T. K., Dean, R. A., and Tyler, B. M. 2007. Expressed sequence tags from *Phytophthora sojae* reveal genes specific to development and infection. Mol. Plant-Microbe Interact. 20:781-793.

Tripathy, S., Pandey, V. N., Fang, B., Salas, F., and Tyler, B. M. 2006. VMD: A community annotation database for oomycetes and microbial

genomes. Nucleic Acids Res. 34:D379-381.
Tyler, B. M. 2002. Molecular basis of recognition between *Phytophthora* pathogens and their hosts. Annu. Rev. Phytopathol. 40:137-167.
Tyler, B. M. 2007. *Phytophthora sojae*: Root rot pathogen of soybean and model oomycete. Mol. Plant Pathol. 8:1-8.
Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., Govers, F., Grunwald, N. J., Huang, W., Ivors, K. L., Jones, R. W., Kamoun, S., Krampis, K., Lamour, K. H., Lee, M. K., McDonald, W. H., Medina, M., Meijer, H. J., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K., Sakihama, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W., Terry, A., Torto-Alalibo, T. A., Win, J., Xu, Z., Zhang, H., Grigoriev, I. V., Rokhsar, D. S., and Boore, J. L. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261-1266.
Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. 1995. Serial analysis of gene expression. Science 270:484-487.
Wang, X. W., Luan, J. B., Li, J. M., Bao, Y. Y., Zhang, C. X., and Liu, S. S. 2010. *De novo* characterization of a whitefly transcriptome and analysis of its gene expression during development. BMC Genomics 11:400.
Xiang, L. X., He, D., Dong, W. R., Zhang, Y. W., and Shao, J. Z. 2010. Deep sequencing-based transcriptome profiling analysis of bacteria-challenged *Lateolabrax japonicus* reveals insight into the immune-relevant genes in marine fish. BMC Genomics 11:472.
Zhou, L., Mideros, S. X., Bao, L., Hanlon, R., Arredondo, F. D., Tripathy, S., Krampis, K., Jerauld, A., Evans, C., St Martin, S. K., Maroof, M. A., Hoeschele, I., Dorrance, A. E., and Tyler, B. M. 2009. Infection and genotype remodel the entire soybean transcriptome. BMC Genomics 10:49.

AUTHOR-RECOMMENDED INTERNET RESOURCES

Joint Genome Institute website: www.jgi.doe.gov
Phytophthora transcriptional database (PTD): phy.njau.edu.cn/ptd