**Running head**

Organization of the Plant Immune Network

**Corresponding author**

Ziding Zhang
State Key Laboratory of Agrobiotechnology
College of Biological Sciences
China Agricultural University
Beijing 100193
China
Tel: +86-10-62734376
E-mail: zidingzhang@cau.edu.cn

**Research Area**

Systems and Synthetic Biology
Revealing Shared and Distinct Gene Network Organization in Arabidopsis Immune Responses by Integrative Analysis

Xiaobao Dong¹, Zhenhong Jiang¹, You-Liang Peng², Ziding Zhang¹

¹State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, China
²State Key Laboratory of Agrobiotechnology and Ministry of Agriculture Key Laboratory of Plant Pathology, China Agricultural University, Beijing 100193, China

Summary
In this work, a machine learning method and modular network analysis were used to characterize the organizing principles of pattern-triggered and effector-triggered immunity in Arabidopsis.
Financial source

This work was supported by grants from the National Natural Science Foundation of China (31271414 and 31471249).

Corresponding author

Ziding Zhang (Email: zidingzhang@cau.edu.cn)
Pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) are two main plant immune responses to counter pathogen invasion. Genome-wide gene network organizing principles leading to quantitative differences between PTI and ETI have remained elusive. We combined an advanced machine learning method and modular network analysis to systematically characterize the organizing principles of Arabidopsis PTI and ETI at three network resolutions. At the single network node/edge level, we ranked genes and gene interactions based on their ability to distinguish immune response from normal growth and successfully identified many immune-related genes associated with PTI and ETI. Topological analysis revealed that the top ranked gene interactions tend to link network modules. At the subnetwork level, we identified a subnetwork shared by PTI and ETI encompassing 1,159 genes and 1,289 interactions. This subnetwork is enriched in interactions linking network modules and is also a hotspot of attack by pathogen effectors. The subnetwork likely represents a core component in the coordination of multiple biological processes to favor defense over development. Finally, we constructed modular network models for PTI and ETI to explain the quantitative differences in the global network architecture. Our results indicate that the defense modules in ETI are organized into relatively independent structures, explaining the robustness of ETI to genetic mutations and effector attacks. Taken together, the multiscale comparisons of PTI and ETI provide a systems biology perspective on plant immunity and emphasize coordination among network modules to establish a robust immune response.
INTRODUCTION

Plants have evolved a sophisticated immune system that enables each cell to monitor every possible destructive invasion by microbe and to mount an appropriate defense response when necessary (Spoel and Dong, 2012). Pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) are two primary immune defense modes in plants (Jones and Dangl, 2006). In PTI, the immune response is triggered when pattern-recognition receptors (PRRs) detect specific molecular patterns from pathogens, also known as pathogen-associated molecular patterns (PAMPs). PAMPs, such as bacterial flagellin, bacterial EF-Tu, lipopolysaccharides (LPSs) and peptidoglycans, are essential components of many pathogens but are lacking in plant cells. Thus, PAMPs are ideal molecular markers for detecting pathogen invasion. Pathogens can secrete virulence proteins called effectors into host cells to subvert the PTI process. Effectors usually mimic the biochemical function of eukaryotic enzymes, such as phosphatases, proteases and ubiquitin ligases, to efficiently block immune signaling pathways at a low dosage (Abramovitch et al., 2006). ETI co-evolved to monitor the presence of pathogen effectors (Chisholm et al., 2006; Jones and Dangl, 2006). In contrast to PAMPs, effectors are directly or indirectly detected by plant resistant (R) proteins in ETI, usually accompanied by a hypersensitive cell death response. After the initiation of PTI or ETI, extensive transcriptional reprogramming occurs, one of the most remarkable phenomena in the plant immune response (Moore et al., 2011).

In both PTI and ETI, the downstream immune response must be tightly controlled by gene networks to balance resource allocation between normal growth and an effective immune response to inhibit pathogen colonization. However, the difference in the downstream immune responses of PTI and ETI remain largely unknown (Dodds and Rathjen, 2010). Genome-wide microarray studies have demonstrated that differences in the Arabidopsis transcriptome between PTI and ETI are largely quantitative (Tao et al., 2003; Truman et al., 2006). It has been proposed that ETI is faster and stronger than PTI and that the signaling network components of PTI and
ETI are similar (Thomma et al., 2011). However, some studies have also suggested that distinct defense regulation mechanisms exist between PTI and ETI (He et al., 2006; Gao et al., 2013), while cross-regulation between PTI and ETI has also been observed (Zhang et al., 2012).

Many immune-related genes involved in PTI or ETI have been identified using genetic screens in combination with biochemistry and molecular biology methods. For example, the Arabidopsis NONEXPRESSOR PR GENES 1 (NPR1) gene is a master regulator of immune response mediated by salicylic acid (SA) (Dong, 2004). NPR1 is required for effective defense in PTI and ETI and is also a crucial molecular switch for the regulation of cell death in ETI.

ENHANCED DISEASE SUSCEPTIBILITY (EDS1) a positive regulator of basal resistance is required for the triggering of ETI by TIR-NB-LRR (Toll-interleukin-1 receptor-nucleotide binding-leucine-rich repeat) type R proteins (Wiermer et al., 2005). Among transcription factors (TFs), WRKY family members play pivotal roles (Pandey and Somssich, 2009). A heat-shock factor-like TF, TBF1, was recently identified as a major TF that controls transcriptional reprogramming in PTI and ETI to favor defense over normal growth (Pajerowska-Mukhtar et al., 2012). Several pioneering studies explored how robust immunity is achieved in PTI and ETI based on these well-known key immune-related genes (Tsuda et al., 2009; Sato et al., 2010; Naseem et al., 2012; Kim et al., 2014). For example, leveraging a dde3/ein2/pad4/sid2-quadruple Arabidopsis mutant, Tsuda et al. (Tsuda et al., 2009) found that immune signaling pathways tend to be synergistic in PTI, but compensatory in ETI. The genes from different signaling pathways tend to positively interact to amplify immune signaling in PTI, whereas in ETI the failure of one signaling pathway can be compensated by other redundant signaling pathways (Tsuda et al., 2009; Dodds and Rathjen, 2010). In 2010, Sato et al. (Sato et al., 2010) performed network modeling and proposed a “sector switching” model to explain the regulatory relationships among 22 Arabidopsis immune-related genes. In this model, negative regulatory relationships among signaling pathways in ETI were derived, that are favorable to the establishment of a robust ETI signaling network against pathogen effectors. For
instance, the perturbation of one pathway caused by effectors can result in the
“switching on” (i.e., high activation) of the other pathways (Sato et al., 2010). To
examine hormone crosstalk in ETI and PTI, network modeling was used to analyze a
phytohormone-centric regulatory network containing 105 nodes (Naseem et al., 2012).
The aforementioned studies have provided invaluable insight into differences in
immune regulation. However, due to the limited network coverage in these studies, a
complete systemic understanding of how immune-related genes coordinate with each
other within gene networks and collectively lead to different PTI and ETI phenotypes
has not been reached (Dodds and Rathjen, 2010).

Because thousands of genes can be differentially expressed after an immune
response is activated, routine analyses based merely on genome-wide gene expression
profiles are challenging. Although simple enrichment analysis can provide abundant
information regarding which pathways are dysregulated, the connections among
dysregulated pathways are usually not apparent and are highly dependent on
interpretation by researchers. As a result, researchers must analyze the identified
pathways individually and may lose the global perspective in understanding how
multiple pathways coordinate with each other and ultimately lead to different PTI and
ETI phenotypes. Breakthroughs in modular cellular network analyses and machine
learning methods can be very useful to dissect behaviors of the immune-related gene
network. Conceptually, many studies have suggested that cellular networks are
modular (Barabasi and Oltvai, 2004; Qi and Ge, 2006) and can be decomposed into
coupled functional units or modules. The genes within one module are usually
co-regulated to perform common functions and contain relatively more dense
interconnections than genes in different modules. The modularity of cellular networks
provides a feasible entry point to simultaneously analyze the complicated behaviors of
many genes. This concept is helpful in understanding plant development (He et al.,
2010; Bassel et al., 2011b), predicting gene functions (Heyndrickx and Vandepoele,
2012; Tzfadia et al., 2012) and comparing stress responses (Shaik and Ramakrishna,
2013). Technically, machine learning algorithms are very good at extracting
information from a large amount of data in an automated way. State-of-the-art
machine learning methods have been used to successfully identify stress-responsive
genes (Ma et al., 2014; Shaik and Ramakrishna, 2014) and development-related gene
associations (Bassel et al., 2011a) from large-scale plant gene expression data.
Recently, Dutkowski and Ideker proposed a new machine learning ranking algorithm
called the network-guided forest (NGF) (Dutkowski and Ideker, 2011). NGF was
developed from Random Forest (RF) (Breiman, 2001), a popular machine learning
algorithm that uses many slightly different decision trees to infer the relationship
between input features and class labels. Because only a randomly selected subset of
training samples and input features are used to build each decision tree, RF is robust
to outliers and can also yield high prediction accuracy (Touw et al., 2013). In the NGF,
the molecular network is further introduced into RF to guide the growth of each
decision tree. For example, by jointly using gene expression states and molecular
network information as input to build the classification model, NGF can identify more
cancer susceptibility genes in cancer stratification compared to conventional methods
that only consider gene expression differences (Dutkowski and Ideker, 2011).

In this study, we developed an integrative network analysis framework that uses a
machine learning algorithm and modular network analysis to investigate the shared
and distinct organization of Arabidopsis gene networks in response to PTI and ETI.
First, we assembled an integrated gene network by combining different types of gene
association data. Using the PTI/ETI gene expression profiles and the integrated gene
network as input, the NGF algorithm was further used to identify key
genesis/interactions involved in the immune response. Moreover, we conducted
multi-resolution network analyses to obtain a global view of the Arabidopsis immune
response. At the individual gene/interaction level, we observed that the gene
interactions involved in the immune response significantly tended to link different
network modules, highlighting the importance of module coordination in plant
defense. At the subnetwork level, we identified a densely connected subnetwork
shared between PTI and ETI and confirmed the biological importance of the shared
subnetwork in the immune response. At the whole network level, we used NGF-based
gene set enrichment analysis (GSEA) (Mootha et al., 2003) to further identify immune
response-related network modules and compared the organization structures of these modules in PTI and ETI. Finally, we have made our results accessible to the scientific community through the development of a user-friendly interactive network web browser.

RESULTS

We designed an integrative multi-step network analysis framework to dissect the complex immune response in Arabidopsis (see Figure 1 for an overview). Herein, we present the assembly of the integrated gene network, the construction of the NGF model, and our major findings at three network levels (individual gene/interaction, subnetwork, and the whole network).

Assembly of an integrated gene network of Arabidopsis

We assembled an integrated gene network by combining four types of gene association data (Figure S1), including (1) experimental PPIs, (2) confirmed protein-DNA binding data, (3) TF and target co-expression relationships inferred from expression profiles under two physiological conditions (development and abiotic stress) and (4) TF and target chromatin co-modification (Marbach et al., 2012) information inferred from genome-wide ChIP-chip experiments of 13 modification types (Table S1). We selected these data because they are representative of and highly complementary to gene regulation relationships (Walhout, 2006; Marbach et al., 2012). To reduce false positives from these high-throughput data, we adopted a series of scoring strategies to remove unreliable gene association data (more details are available in the Materials and Methods, Figures S1-S5).

Regarding the experimental PPI data, the reliability of each PPI was measured using the scoring scheme introduced by Lage et al. (Lage et al., 2007), in which the network topology information surrounding the interaction, the number of publications supporting the interaction, and the experimental scale of PPI determination are combined to compute a probabilistic confidence score. 77 PPIs with confidence scores...
< 0.20 were removed (Figures S1 and S2). Because gene association pairs tend to be
involved in a common biological process, the gene ontology (GO) annotation
similarity of gene pairs can also provide indirect evidence of the reliability of PPIs.
The retained PPI data yielded an average GO annotation similarity score of 0.158,
which is significantly higher than that of a random network (Figure S3). The average
GO annotation similarity of retained PPIs can stably increase with the improvement of
confidence score cutoff (Figure S4). However, the number of retained PPIs also
decreases considerably when improving the cutoff (Figure S4). Therefore, it is
difficult to select an optimal cutoff. Considering that experimental PPI data are
relatively insufficient in Arabidopsis, we only used the cutoff of 0.20 to remove PPIs
with very low reliability scores.

We inferred three gene networks from gene expression or chromatin modification
data. First, two gene expression data sets collected during plant development (Schmid
et al., 2005) and the abiotic stress response (Kilian et al., 2007) were used to construct
two co-expression networks. The experimental conditions of the two datasets
represent two major themes in the life of plants (development and adaptation to the
environment). The diverse treatments and rich data (> 200 profiles for each dataset)
also make these datasets suitable for inferring co-expression relationships. Second, we
constructed a gene chromatin co-modification network because particular chromatin
states are preferentially associated with gene classes involved in development and the
response to pathogens (Luo and Lam, 2010). To construct the network, we developed
a modification pattern similarity measurement to assess whether two genes have
similar chromatin states. Using the average GO annotation similarity score of the
filtered PPI data as a reference, we further filtered the gene interactions in the three
networks to ensure that the average GO annotation similarity score of the remaining
interactions in these three networks was equal to 0.158. Finally, we only retained
inferred interactions in which at least one gene was a TF.

We did not filter the experimentally confirmed protein-DNA binding data. The
above refinement yielded five different networks, and the union set of these five
networks was computed to generate an integrated gene network encompassing ~50%
of the Arabidopsis genome [11,155 nodes (including 1,090 TFs) and 64,281 edges].
To combine the network data with immune response expression profiles, we further
removed from the network those genes that were not designed on the ATH1
microarray. The retained network (AraONE, Table S2) used by NGF included 9,862
genes and 57,487 interactions. The overlaps among the five networks are presented in
Figure S5.

Compared to the pure PPI network, AraONE has better genome coverage and
reflects multiple gene associations in Arabidopsis. Although false positive interactions
may inevitably be present in AraONE inevitably, our benchmark analysis
demonstrates that the average GO annotation similarity of AraONE is significantly
higher than that of a random gene interaction network (Figure S3). Therefore, we
argue that AraONE is suitable for further network analyses.

**Ranking immune-related genes and their interactions using the network-guided forest algorithm**
NGF uses many slightly different network-guided decision trees that combine
gene expression and gene networks simultaneously to yield an accurate classification
model. The NGF tree structure is suitable for modeling complex gene expression
combinatorial logic, such as synergistic or antagonistic effects. NGF assigns an
importance score (IS) for each gene/interaction according to its contribution to
improving the model classification accuracy.

We constructed two NGF models to classify PTI versus the control (PTI group)
and ETI versus the control (ETI group). To train the NGF models, we compiled three
Arabidopsis gene mRNA microarray datasets corresponding to three different immune
response scenarios (PTI, ETI and control) from multiple public databases (see the
Materials and Methods for additional details about the microarray datasets, Table S3;
and model building procedures). To evaluate and refine the NGF results, we
conducted a series of computational experiments.

First, we assessed the prediction accuracy of NGF through a 10-fold
cross-validation test. Accuracy (Acc), which represents the proportion of correctly
predicted cases under a default cutoff, and the area under the receiver operating
characteristics curve (AUC), which reflects the average accuracy of the classification
models under different cutoffs, were used. As shown in Table 1, NGF was highly
accurate (Acc ≈ 0.86 and AUC ≈ 0.94) for both the PTI and ETI groups. The
performance was also validated on two independent datasets (Table S4). Considering
the randomness of NGF, we further examined the reproducibility of the NGF ranking
results. Accordingly, we built 10 more NGF models with the same settings for the PTI
and ETI groups, respectively, and performed a pairwise comparison between the
models in the same group. More than 80% of the genes were reproducibly identified
as significantly important immune-related genes at multiple scales in both the PTI and
ETI groups (Figure S6). For gene interactions, more than 80% and approximately 70%
were reproducible for the PTI and ETI groups, respectively (Figure S6). The relatively
low reproducibility of the ETI group may be attributable to less training sample data.
However, given the large number of gene interactions in AraONE, the NGF
interaction ranking results are very stable.

Finally, we permuted the training data labels and re-built 100 null-hypothesis
models. We compared real ISs with the ISs from null-hypothesis by \( t \)-test. Only those
genes and interactions with statistically significant ISs were subjected to further
analysis (see the Material and Methods). Based on these computational tests and the
statistical analysis, we are confident of the reliability of the NGF ranking results.

For PTI and ETI group, there were 1,856 and 1,843 genes, respectively, were
identified as significantly important (Bonferroni correction, \( t \)-test, \( p \)-value < 0.05;
Table S5). We list the top 30 genes with the highest ISs in Tables 2 and 3 (see also
Table S6) for further analyses. Of the top 30 genes, our approach successfully
identified 12 and 14 known immune-related genes for PTI and ETI, respectively. We
compared the NGF gene ranking results with differentially expressed genes (DEGs).
We defined the differential expression change of a gene as the absolute value of the
difference of its average expression levels under two conditions (PTI/ETI vs. control).
Many of the top 30 DEGs are also involved in plant immunity (Table S7). We
observed that the top-ranked genes based on NGF are prone to differential expression
in either PTI or ETI (Figures 2A and B). However, highly differential expression of a
gene is not necessary to obtain a high rank in the NGF results. As shown in Figures
2C and D, only two genes of the top 30 DEGs are ranked as the top 30 significantly
important genes in the NGF. For both the PTI and ETI groups, most of the top 30
DEGs have a low network degree ($\leq 2$) in AraONE (Table S7). This low degree might
make these DEGs more difficult to be introduced to decision trees. By contrast, some
immune-related genes with moderate expression changes were assigned a high rank,
such as the calcium-dependent protein kinase (CPK) gene $CPK1$ (Coca and San
Segundo, 2010) in the PTI group and $NPR1$ in the ETI group, which might be
neglected in traditional gene differential expression analyses. Closer inspection of
these two genes in the network revealed that $CPK1$ and $NPR1$ are highly connected
and that many of their neighbors are DEGs. In fact, these DEGs also include known
immune-related genes, such as the $NPR1$ neighbors $GRX480$, $NIMIN1$, $TGA4$ and
$WRKY25$, and the $CPK1$ neighbor $LHT1$ (Figure S7). Although $CPK1$ and $NPR1$
exhibit only moderate changes in expression, the NGF algorithm combines their
expression changes and network topologic features to identify them as
immune-related genes among thousands of candidates.

In addition to assigning an IS for each gene, one important feature of the NGF is
that it can also yield an IS for each interaction in the network. In total, 5,466 and
6,342 significantly important interactions ($p$-value $< 0.05$, Bonferroni correction,$
t$-test; Table S5) corresponding to 3,304 and 3,629 genes were identified by NGF from
the PTI and ETI group, respectively. Because most functional annotation systems are
gene-centric, the biological meaning of the identified interactions is difficult to
interpret. Since previous studies have indicated that interactions connecting different
cellular network modules tend to play an important role in specific biological
processes (He et al., 2010; Bassel et al., 2011b), we also examined the distribution of
the identified interactions in the context of network modules. We first applied the
Markov Cluster algorithm (MCL) (Enright et al., 2002) to detect densely connected
network modules in AraONE. A total of 633 modules (Table S8) were identified, and
47% of which contained more than five genes. Second, the inter-module to
intra-module interaction ratio (MIR) of significantly important interactions was calculated. Interactions in which the two genes belong to the same module were denoted intra-module interactions; interactions in which the two genes belong to different modules were denoted inter-module interactions. The MIR of a set of interactions was then defined as the number of inter-module interactions divided by the number of intra-module interactions. As a control, we randomly sampled the same number of interactions identified by NGF from AraONE. This procedure was repeated 1000 times. The MIRs for the PTI and ETI groups were much larger than those of the corresponding control group (0.667 vs. 0.344 and 0.587 vs. 0.344, respectively; p-value < 10^{-100}, t-test; Figures 2E and F), and the MIR values increased with a more stringent IS cutoff (Figures 2E and F). This observation indicates that inter-module interactions were assigned relatively high ISs, suggesting that genes that distinguishing the immune response and control samples usually came from multiple network modules instead of being enriched within one module. This observation may partly reflect the widespread gene expression changes in the plant immune response, which involves multiple biological processes (Wang et al., 2011). The plant immunity-related functional roles of the significantly important inter-module interactions (Table S5) merit further experimental validation.

**The shared subnetwork between PTI and ETI is enriched with effector targets**

To explore the shared network components between PTI and ETI, we computed the intersection of significantly important interactions in the two processes. The resulting subnetwork includes 1,156 genes and 1,289 interactions (Figure 3A), of which 43.9% (566/1,289), 1.5% (20/1,289), 0.5% (6/1,289), 52.0% (671/1,289) and 1.2% (16/1,289) were exclusively from PPI data, protein-DNA binding data, gene chromatin co-modification data, developmental gene co-expression data and abiotic response gene co-expression data, respectively. Nineteen of the 181 TFs in the subnetwork belong to the WRKY family. Consistent with the MIR analysis, many interactions in this subnetwork tended to link modules (observed: 495/1289, expected: 312/1289, p-value=2.20×10^{-16}, t-test). Rather than forming many interrupted small
subgraphs, most genes are connected to a large subgraph encompassing 946 genes and 1,149 interactions (Figure S8). We used a new gene function visualization method called Enrichment Map (Merico et al., 2010) to characterize the functional distribution of genes in the shared subnetwork (Figure 3B). In the Enrichment Map, each node represents a gene set corresponding to a GO term enriched in the subnetwork, and each weighted edge represents the overlap between two gene sets. We observed that stress response and development were two major themes on the map, and regulations of defense and development were linked by the term “negative regulation of cellular process”. These two themes remained prominent even after the interactions inferred from the co-expression data were removed (Figure S9).

We further analyzed the distribution of SA response genes and type-III effector target genes in this subnetwork. SA is a primary plant hormone in resistance to infection by biotrophic pathogens and participates in both PTI and ETI. To examine the distribution of SA response genes in this subnetwork, we used an independent dataset generated by Wang et al., in which 2,288 genes were identified using a genome-wide microarray after 24h of SA treatment on Arabidopsis (Wang et al., 2006). Note that the microarray data were not used to construct NGF models. A total of 1,303 SA response genes could be mapped on AraONE, 322 of which were included in the shared subnetwork (p-value < 10^{-100}, hypergeometric test), indicating that this shared subnetwork is significantly enriched in the SA response. Moreover, we also downloaded 165 putative effector targets identified by Mukhtar et al. (Mukhtar et al., 2011) and mapped them on AraONE. These targets were defined as Arabidopsis proteins that physically interact with pathogen effectors. A total of 128 target genes could be mapped on AraONE, 49 of which were observed in the shared subnetwork (Table S9; Data S1), indicating that effector targets are also enriched in the subnetwork (p-value=6.44×10^{-15}, hypergeometric test). Taken together, we observed that the subnetwork is highly enriched in the SA response and also tend to include pathogen effector targets. Therefore, this subnetwork may represent an important part of the plant immune network shared by PTI and ETI. However, the
complete functional role of this subnetwork in plant immunity remains to be
established through additional experiments.

Modular network models of PTI and ETI

To delineate the global architecture of the gene networks activated in PTI and ETI,
we applied GSEA to identify immune response-related network modules and further
constructed two modular network models. GSEA is a nonparametric rank-based
method in which information from all members of a gene set is combined as a whole
to improve the signal-to-noise ratio. One prominent advantage of GSEA is that it can
detect modest but significantly coordinated changes in a gene set that may be missed
by individual gene-based analyses, such as a $t$-test (Mootha et al., 2003). GSEA was
originally introduced as a method to identify dysregulated pathways in microarray
experiments but can also be used with other data types with the same structure as gene
expression profiles. Here, we performed GSEA by replacing gene expression changes
with NGF scores (i.e., NGF-based GSEA). Using the 100 null-hypothesis models as
the background, we identified network modules significantly enriched in genes with
high ISs in the PTI or ETI groups (see the Materials and Methods for details). A total
of 62 and 52 modules were identified by NGF-based GSEA in the PTI and ETI groups
(permutation test, empirical $p$-value < 0.05; Table S10), respectively, and 25 modules
were common between the two groups. We further conducted GO functional
annotations of these modules using BiNGO (Maere et al., 2005). For convenience, we
manually classified these modules into five functional categories according to their
GO term annotations: signal transduction, transcription regulation, hormone, defense
response and growth or development. Those modules without any category
assignment or GO enrichment were omitted from further analyses. Finally, we
obtained 18 PTI-specific modules, 14 ETI-specific modules and 19 common modules
(Table S11).

Moreover, we computed average logic state scores for the genes in each module
and interactions between any two connected modules to characterize the combination
patterns of the gene expression changes in immune response. In contrast to ISs, which
were used to represent the usefulness of a gene or interaction in successfully distinguishing the immune response and control conditions, logic state scores reflect the direction of gene expression change in the immune response relative to control conditions. The logic state score of a gene was derived from the decision rules learned by NGF models. A positive/negative logic state score for a gene indicates that the gene state is “up-regulation”/”down-regulation” for the majority of decision rules ended as immune response. Its absolute value of this score is proportional to the number of decision rules supporting the gene state. Moreover, we also computed the logic state scores of interactions that link different modules. For an interaction, a positive logic state score means the logic states of two interacting genes are identical, while a negative interaction logic score denoted different gene logic states between two interacting genes (see the Materials and Methods for more details).

The resulting modular network modes of PTI and ETI are shown in Figure 4. By observing these two modular network models, we found that protein phosphorylation modules were the most highly connected in both the PTI and ETI models, suggesting that protein phosphorylation is a core biological process linking other functional modules in the plant immune response.

Closer scrutiny of the substructures of these two models that only include the modules of signal transduction and defense response categories yielded an unexpected finding. In PTI, defense response modules appear to be densely linked by a central protein phosphorylation module, resulting in a cohesive clique with prevalent positive logic interactions (Figure 5A). By contrast, defense response modules in ETI are divided into multiple relatively independent parts and contain only a few links through the central protein phosphorylation module (Figure 5B). The advantage of such module organization in ETI is that the perturbation of one module affects only the activity of the neighboring modules instead of spreading to the all defense response modules. As proposed by Tsuda et al. based on genetic analysis, the signaling pathways in PTI tend to interact synergistically, whereas the signaling pathways in ETI tend to be compensatory. Our findings are in line with the observations of Tsuda et al. (Tsuda et al., 2009; Tsuda and Katagiri, 2010).
Module M63 suggests a potential role of chromatin dynamics in ETI

We observed a highly connected module (M63) specific to the ETI model with an NGF-based GSEA empirical \( p \)-value of 2.77\( \times 10^{-3} \) (permutation test, empirical \( p \)-value = 0.730 in PTI). M63 contains 1,030 genes and is the largest module in AraONE. Although the most significant GO term annotated by BiNGO was “gene expression”, other significant terms such as “DNA conformation change”, “nucleosome assembly”, “ribonucleoprotein complex biogenesis”, “chromatin modification”, “DNA replication” and “DNA repair” were also obtained. To summarize the biological themes in M63, we submitted M63 to the DAVID webserver (Huang et al., 2008) to cluster redundant GO annotations. Most of the enriched themes were related to chromatin configuration (Figure S10), a highly relevant process for gene expression and that is an intrinsic component of plant-bacteria interactions (Ma et al., 2011). Considering the large size of M63, we further decomposed this module into four sub-modules (Figure S11). Most of the genes in the largest sub-module (M63.1, size=358) were connected to \( \text{NAP57} \), \( \text{ATWHY2} \), \( \text{HDA3} \) and \( \text{HD2C} \). \( \text{NAP57} \) a component of telomerase ribonucleoprotein and is required for telomere maintenance (Kannan et al., 2008). \( \text{ATWHY2} \) is a mitochondrial protein that is involved in organelle genome repair (Cappadocia et al., 2010). \( \text{HDA3} \) and \( \text{HD2C} \) are two histone deacetylases that are involved in development and the stress response (Wu et al., 2000; Luo et al., 2012). Consistent with the NGF-based GSEA analysis, the four genes were distributed at the top of the NGF rank list. \( \text{NAP57} \) and \( \text{ATWHY2} \) were included in the top 30 genes by NGF in the ETI group (Table 2). \( \text{HDA3} \) and \( \text{HD2C} \) were ranked as the 63rd and 47th genes. However, the functions of these genes in plant immunity remain unclear.

We further compared the average closeness centrality (CC) of the genes in M63 with the genes in the whole network. CC is a network topological parameter that measures if a node is near the center of the entire network. The average M63 CC values were significantly higher than that of the whole network (\( p \)-value =9.41\( \times 10^{-324} \), \( t \)-test, see Figure S12). Therefore, M63 gene activities could quickly influence the entire network. Moreover, we observed a recently identified epigenetic regulator in M63, \( \text{ELP2} \), that is essential for rapid transcriptional reprogramming in ETI (Wang et
al., 2013). Taken together, M63 implies a potential role of chromatin dynamics in ETI, although further validation is required.

**A web tool to interactively explore modular network models**

To facilitate the exploration of these network models by the scientific community, we developed a user-friendly interactive network web browser (http://systbio.cau.edu.cn/pinet/home.php) based on the network visualization framework sigma.js (http://sigmajs.org). This portal includes detailed gene composition information for each module from two modular network models as well as the shared subnetwork identified using NGF.

**DISCUSSION**

PTI and ETI are two major types of plant immunity; the dissection of their shared and distinct aspects is valuable for further understanding the evolution of plant immunity and can provide important insight for breeding new generations of disease-resistant crops. In this study, we used a network-based analysis framework and the machine learning algorithm NGF to integrate heterogeneous but complementary high-throughput genomics data to gain a systems-level understanding of how different parts of the plant immune network coordinate with each other to lead to specific phenotypes.

**NGF can effectively combine gene network and expression cues to pinpoint immune-related genes**

We used NGF to assess the importance of genes and interactions in the specific immune response. By converting the selection of immune-related genes into a task to classify the conditions of the expression profiles, NGF automatically identified immune-related genes and it reveals advantages over some existing methods used in plant systems biology. Compared to other differential expression analyses based on machine learning methods (Shaik and Ramakrishna, 2014), the major innovation of NGF is that it incorporates the gene network as a priori knowledge to effectively
narrow the hypothesis space of candidate genes. In contrast to network-based methods that rely on known functional genes to discover new candidates (Lee et al., 2010; Ma et al., 2014), NGF can be regarded as a de novo gene discovery algorithm, which is particularly useful when only few genes involved in the biological process of interest are known.

Indeed, many significantly important genes identified by NGF are known immune-related genes. For example, multiple WRKY family TFs were considered significantly important genes by the NGF in the PTI group (Table 2), including \textit{WRKY53}, \textit{WRKY17}, \textit{WRKY33} and \textit{WRKY25}. \textit{WRKY53} was first identified as a senescence-associated gene (SAG) and subsequently screened out as a member of the systemic acquired resistance (SAR) signaling network (Hinderhofer and Zentgraf, 2001; Wang et al., 2006). This gene can interact with the jasmonic acid (JA)-inducible protein ESR to mediate negative crosstalk between pathogen resistance and senescence (Miao and Zentgraf, 2007). Recent genetic analysis has suggested that \textit{WRKY53} positively regulates basal resistance by coordinating with \textit{WRKY70} and \textit{WRKY46} (Hu et al., 2012). By contrast, \textit{WRKY17}, \textit{WRKY33} and \textit{WRKY25} are negative regulators of the defense response to \textit{Pseudomonas syringae} (Journot-Catalino et al., 2006; Zheng et al., 2006; Zheng et al., 2007). \textit{CPK1} and \textit{CPK4} are also ranked in the top 30 significantly important immune-related genes in the PTI group. A loss-of-function mutation of \textit{CPK1} increases the susceptibility of \textit{Arabidopsis} to pathogen infection (Coca and San Segundo, 2010). \textit{CPK4} has been identified as a key positive component in initial PTI signaling to flg22. This protein directly phosphorylates WRKY8/28/48 (Gao et al., 2013). In the ETI group, \textit{NPR1} and \textit{EDS1} were identified by NGF as the 20th and 25th most significantly important genes, respectively. \textit{WRKY70} a convergence node for JA-mediated and SA-mediated signals in plant defense (Li et al., 2004), is also included in the list for the ETI group.

Several of the top-ranked genes that have not previously been related to the plant immune response may be candidates for further experimental validation. For example, \textit{LOV1} (At2g02450) was ranked as one of the top 30 significantly important immune-related genes in both the PTI and ETI groups. \textit{LOV1} belongs to the NAC
transcription factor family, of which some members have been found to be involved in the defense response to pathogen infection (Bu et al., 2008). *LOV1* is a master regulator of flowering and cold response (Yoo et al., 2007), and its overexpression in switchgrass induces alterations lignin content and cell wall composition (Xu et al., 2012). Plant development and innate immunity are closely linked, and previous studies have uncovered several genes that can regulate both flowering time and plant immunity, including *WIN3, NPR1* and *PUB13* (Wang et al., 2011; Li et al., 2012). Therefore, we infer that *LOV1* is an immune-related candidate gene that warrants further characterization. Taken together, these results demonstrate the power of NGF for identifying key immune-related genes. Given the increasing availability of genome, proteome and interactome data, we expect that NGF will be a useful tool for modern plant research.

**Identification of a shared subnetwork between PTI and ETI**

By comparing the significantly important gene interactions between PTI and ETI, we identified a shared subnetwork enriched in effector targets. The subnetwork spans the extracellular region to the nucleus (Figure 3A) and includes many key immune-related genes, such as *NPR1, TBF1, PEN3, BAK1, RIN4* and *HBI1*. *PEN3* is an ATP binding cassette transporter that participates in defense against fungal and oomycete pathogens (Clay et al., 2009). *BAK1* is a master positive regulator of PTI (Schwessinger et al., 2011). *RIN4* is a known target of multiple effectors (Jones and Dangl, 2006). *HBI1* is a bHLH TF that mediates the trade-off between plant growth and the immune response (Fan et al., 2014). Functional enrichment analysis indicated that development and defense are the most prominent themes in this subnetwork. These two themes were also observed in a conserved *Arabidopsis* PPI network targeted by effectors of pathogens from three kingdoms of life (Weßling et al., 2014). This shared subnetwork between PTI and ETI might contribute to the coordination of different biological processes by linking together relatively separated network modules to promote plants change their state from development to defense. In addition, the enrichment of pathogen effector targets implies pathogens might exploit
genes in the subnetwork to impede plant immunity. In addition to TBF1 and HBI1, we speculate that additional genes in this subnetwork might function as regulators of the trade-off between development and defense.

A relatively independent organization of defense modules distinguishes ETI from PTI

Two modular network models were constructed to compare the global network architecture of PTI and ETI. Protein phosphorylation was observed as a core process linking other network modules in both the PTI and ETI models. Reversible protein phosphorylation is a crucial mechanism for regulating protein activity and is a basic step in protein kinase-mediated signal transduction. Signaling networks mediated by protein kinases have been characterized as essential components of plant innate immunity (Tena et al., 2011). For example, receptor protein kinases (RPKs) are the main PRRs sensing diverse PAMPs in PTI. Mitogen-activated protein kinase (MAPK) cascades were identified as defense response-related signaling pathways that directly connect upstream immune signaling from RPKs and downstream transcriptional reprogramming. Moreover, the CPK signaling pathway was also recently discovered as another convergence point of the upstream plant immune response and regulates ETI-related programmed cell death (PCD) (Gao et al., 2013). In addition to regulating the plant immune response, protein phosphorylation signaling networks participate widely in other plant physiological processes, including response to light, abiotic stress, hormone stimulus and nutrient deprivation (Stone and Walker, 1995). Therefore, it is reasonable that protein phosphorylation nodes are located in the center of these two immune response models and link most of the modules from five functional categories (Figure 4).

However, we observed defense response modules that were organized into a relatively independent network structure in ETI in contrast to the cohesive structure observed in PTI (Figure 5). The distinct defense module organization in PTI and ETI may be attributable to different demands in plant-pathogen coevolution (Katagiri and Tsuda, 2010). PTI is triggered by conserved PAMPs, but many non-pathogenic
microbes also carry PAMPs. Plants must carefully balance the induction of strong
immune responses to clear pathogens and the prevention of undue immune responses
triggered by non-pathogenic microbes. The interdependent relationships among
defense modules in PTI would facilitate the gradual establishment of a strong immune
response when there are persistent PAMP signals, which could prevent the side effects
of frequently activated immune responses triggered by invasion signals from low-risk
pathogen. By contrast, the recognition of effectors by R proteins is a high-risk signal
that requires plants to quickly adopt defense responses. The relatively independent
relationships among defense modules in ETI could be the result of evolutionary
demands for a rapid, robust defense response.

Limitations and future work
Our results are only based on currently available data and must be interpreted
with caution. First, our approach is limited by the availability and quality of the
molecular network. For example, we used an *ad hoc* cutoff to filter experimental PPIs.
Many gene interactions were inferred from genomic data without confirmation, and
only approximately half of *Arabidopsis* coding genes are covered by our analysis.
Imperfect data quality and insufficient data coverage can result in errors. For instance,
interactions in AraONE most likely include both direct and indirect gene regulations
because many of these interactions were inferred from omics data instead of direct
and detailed biochemical evidence. Thus, it is very difficult to judge whether the
network organizations observed here are true characteristics of the underlying gene
regulatory architecture or reflections of gene functional associations in the immune
response. Second, mRNA microarray data provide no information of the regulation of
immune-related genes by post-translational modifications. We believe that these
problems will be alleviated gradually with the development of more reliable
computational algorithms (Lee et al., 2010) and high-throughput experimental
methods (Jones et al., 2014; Lumba et al., 2014). Finally, NGF only distinguishes
plant immune responses as two conditions (on/off). However, the plant immune
response is a highly dynamic process. Dynamic behavioral information about the
immune regulatory network that could be crucial in discriminating PTI and ETI is inevitably lost in a simple binary discrimination of the plant immune response (Katagiri and Tsuda, 2010). Integrating time-course expression data with module analysis will further decipher the regulators and the core regulatory network that drives different PTI and ETI expression kinetics.

**MATERIALS AND METHODS**

**Data collection and processing**

To assemble the integrated gene network, we (1) compiled an experimental PPI network from BioGrid, IntAct and TAIR that included 6,640 proteins and 16,797 interactions; (2) downloaded protein-DNA interactions annotated as “confirmed” from AGRIS (http://arabidopsis.med.ohio-state.edu/) covering 720 genes and 762 interactions; (3) downloaded 237 and 298 preprocessed gene expression profiles from the Weigel World website (http://www.weigelworld.org/), which are related to Arabidopsis development and the abiotic stress response, respectively; and (4) collected 13 types of chromatin modification data from ChIP-chip experiments (see Supplemental Table S1).

The ChIP-chip data for 13 chromatin modification types were processed using TileMap (Ji and Wong, 2005) to detect modification regions in the genome. Similar to the work of Marbach et al. (Marbach et al., 2012), chromatin modification information was mapped to five regions of a gene to characterize the chromatin modification status. These five regions were the 1kb upstream region of the transcriptional start site (TSS), the 5’UTR, the coding region, the 3’UTR and the 1kb downstream region of the 3’UTR. For each type of chromatin modification, we constructed a 5-dimensional binary vector to determine if the above five regions of a gene were modified [34]. For example, [0, 1, 1, 0, 0] denotes that the 5’UTR and coding region of a gene were modified by a specific chromatin modification. The binary vectors of the 13 types of modification were concatenated into a 65-dimensional vector to represent the chromatin status of a gene. Genome
information was downloaded from TAIR (http://www.arabidopsis.org).

To implement the NGF algorithm, *Arabidopsis* gene mRNA microarray data
(Affymetrix *Arabidopsis* ATH1 Genome Array) corresponding to three different
immune response scenarios (PTI, ETI and control) were downloaded from the NCBI
Gene Expression Omnibus (GEO), EBI ArrayExpress and TAIR. Regarding the PTI
data, 69 gene expression profiles were collected from *Arabidopsis* after different
bacterial PAMPs treatments, including flg22, EF-Tu, HrpZ, LPS and *Pseudomonas*
syringae pv. DC3000 with a deficient type-III secretion system. Regarding the ETI
data, we collected 36 gene expression profiles of *Arabidopsis* in response to
*Pseudomonas syringae* species carrying the Avr gene (*AvrRpm1* or *AvrRpt2*), which
can elicit ETI in Col-0 type *Arabidopsis*. The control data included 65 gene
expression profiles gathered from the corresponding control experiments for the
above treatments. More detailed information can be found in Supplemental Table S3.

Raw data in CEL format were processed using the robust multi-chip average (RMA)
method with the affy package from Bioconductor (http://www.bioconductor.org) and
further transformed into z-scores to remove batch effects.

**Filtering PPI network**

The reliability of a PPI was measured using the scoring scheme of Lage et al.
(Lage et al., 2007). First, the network topology information surrounding the
interaction was used to compute a raw score (*RS*) between \((-\infty, 0]\).

\[
RS(g_1, g_2) = -\ln\left(\frac{(NS_1 + 1)(NS_2 + 1)}{N} \right)
\]

where *NS*₁ and *NS*₂ denote the number of non-shared protein interaction partners of *g₁*
and *g₂*, respectively. The smaller number of shared PPI partners results a small value
of *RS*. Second, the number of publications supporting the interaction, and the
experimental scale of the PPI determinations were combined to post-process the *RS* as
follows:

\[
score(g_1, g_2) = \frac{RS}{\sum_{i=1}^{N} 1/\ln T_i}
\]
\[
normalized_score \left( g_1, g_2 \right) = \frac{score(g_1, g_2) - score_{\text{min}}}{score_{\text{max}} - score_{\text{min}}}
\]

where \( N \) is the number of different PubMed IDs corresponding to publications in which this interaction is reported. \( T_i \) (throughput) is the number of PPIs reported in publication \( i \). Thus, if an interaction was only reported in a high-throughput experiment, \( score(g_1, g_2) \) will be small. Finally, this score was normalized to \((0, 1]\) using Equation 3. To select a suitable cutoff for removing unreliable PPIs, we plotted the distribution curve of the remaining PPI numbers at different cutoffs (Figure S2). A visual determination of the “elbow” was made and the cutoff of 0.2 was chosen to remove PPIs with very low reliability scores.

**Computing GO term annotation similarity**

First, the similarity of two GO terms was measured using the protocol developed by Bradford et al. (Bradford et al., 2010), in which both the directed acyclic graph (DAG) structure of GO annotations and functional diversity of a gene were considered. The occurrence frequency \( p(c) \) of a GO term was calculated as the number of *Arabidopsis* genes annotated with that term divided by the total number of genes annotated with any GO term. The probability of minimum of subsume \( p_{\text{ms}}(c_1, c_2) \) between the two terms \( c_1 \) and \( c_2 \) was defined as the minimum of \( p(c) \) in all possible shared parent GO terms. Then, the similarity of the two terms \( c_1 \) and \( c_2 \) was measured as \(-\ln p_{\text{ms}}\). Second, the GO annotation similarity of \( g_1 \) and \( g_2 \) was measured according to the similarity of the GO terms. The \( m \) terms associated with \( g_1 \) and \( n \) terms associated with \( g_2 \) were used to construct an \( m \times n \) similarity matrix. The sum of the maximum term similarity scores from each line and each column were divided by \( m+n \) to give a GO annotation similarity score between \( g_1 \) and \( g_2 \). In this process, we only considered the GO term “biological_process” and its child terms. Note that GO annotations were downloaded from TAIR 10 (http://www.arabidopsis.org/), and only those terms with the evidence codes “IDA”, “IPI”, “IMP”, “IGI”, “RCA”, “TAS”, “ISS”, “ISO”, “ISA”, and “ISM” were considered validated annotations.
Construction of three gene networks based on gene expression profiles and gene chromatin modification data

We inferred three gene interaction networks (Figure S1) using gene co-expression patterns in Arabidopsis development and the abiotic stress response, and the gene co-modification patterns of the 13 chromatin modification types.

The squared Spearman correlation coefficient was used to measure the correlation of gene co-expression patterns because this coefficient has been widely used and showed more accurate performance than Pearson correlation coefficient in inferring TF-target relationships (Ma and Wang, 2012). To minimize the potential possibility of introducing bias to the NGF algorithm, we avoided using immune response-related expression profiles to construct the co-expression network.

We used a weighted binary vector matching coefficient to measure the similarity of chromatin modification between two genes as follows:

\[
\text{comodi}(g_1, g_2) = \sum_{m \in M_{\text{all}}} \frac{w(m)}{\sum_{m \in M_{\text{all}}} w(m)}
\]

where \( m \) denotes a specific modification in the binary chromatin modification vector; \( M_{\text{all}} \) represents all 65 of the different modifications in the vector; \( w(m) \) is the weight of \( m \), which is calculated using Equation 5. \( N \) is the number of genes in the genome and \( n_m \) is the number of genes with modification \( m \).

Moreover, the cutoffs 0.70, 0.80 and 0.44 were used to filter the interactions inferred from gene co-expression in development, gene co-expression in the abiotic stress response and chromatin co-modification, respectively. Thus, the average GO annotation similarity score of each filtered network was identical to that of the retained PPI network. Finally, we only retained the co-expression/co-modification pairs in which at least one gene was a TF according to the TF annotations of AGRIS.

Network-guided forest algorithm

The NGF uses the gene expression profiles from two different conditions and the
gene interaction network as input to train a classification model that can predict the phenotype of a given expression profile. NGF assigns a score called IS to each gene/interaction according to its contribution to improve the classification accuracy. The complete methodology is detailed in the work of Dutkowski and Ideker (Dutkowski and Ideker, 2011). The gene expression profiles of PTI, the control and AraONE were used to train a PTI-related NGF model, whereas the gene expression profiles of ETI, the control and AraONE were used to train an ETI-related NGF model. To generate one decision tree, a subset of all expression profiles was first randomly sampled with replacement to produce a new expression profile dataset with the same size as the original one. Next, a best predictive gene was selected from a randomly sampled gene subset (size = \( \sqrt{N_g} \)) as a root. \( N_g \) is the number of genes in AraONE. To guarantee sufficient candidate genes for use in the next step, only genes with degree \( \geq 5 \) were used as root candidates. In the tree growth, the decision tree iteratively partitioned gene expression profiles according to the expression level of the gene (decision gene). In each step of tree growth, decision gene candidates were defined as network neighbors of the genes that had been included in the decision path, and a new decision gene was further selected as that with the highest \( IS(g) \) (see Equation 6 below). The tree growth was terminated until the statistical predictive power (\( IS(g) < 0.02 \)) no longer improved when a new gene was added. A total of 10,000 decision trees were built to construct the NGF model. We implemented NGF using custom Python scripts based on NetworkX (http://networkx.github.io/).

In the original NGF algorithm, the ISs of genes and interactions are measured using the average reduction of tree accuracy on out-of-bag samples. We modified the original algorithm by using a faster method to calculate the ISs. In brief, we used a reduction in the output sample variance due to the node (gene) split in the decision tree to measure the importance of the gene, which is defined as follows:

\[
IS(g) = var(T) - \frac{n_U}{N_T} var(U) - \frac{n_D}{N_T} var(D)
\]
\[ \text{var}(S) = \text{Gini}(S) = 1 - \left( \frac{n_p}{N_S} \right)^2 - \left( \frac{n_N}{N_S} \right)^2 \]  

where \( \text{var}(T) \) is the variance of all of the samples as tested by split node \( g \); \( \text{var}(U) \) is the variance of the sample subset \( U \) in which the expression level of split node \( g \) exceeds the threshold (up-regulated); and \( \text{var}(D) \) is the variance of sample subset \( D \) in which the expression level of node \( g \) is lower than or equal to the threshold (down-regulated). \( N_T, n_U \) and \( n_D \) are the sizes of the corresponding sample sets. The Gini index was used to measure the sample variance as defined by Equation 7, in which \( n_P \) and \( n_N \) represent the number of positive and negative samples in sample set \( S \), respectively.

The ISs of the genes were further used to reversely infer the importance of interactions. When gene \( g \) was added to the decision tree, all \( N \) interactions between \( g \) and its parent nodes in the decision tree were assigned an IS equal to \( IS(g)/N \). If a gene or interaction was used more than once in a decision tree, all of its ISs were summed. The ISs from all 10,000 decision trees in the NGF model were further averaged to produce the final IS.

To estimate the statistical significance of ISs for genes and gene interactions, we compared real ISs with those inferred from random conditions and assigned a \( p \)-value for each real IS. For example, in the PTI group, we first randomly permuted the condition labels (i.e., PTI and control) of 69 PTI microarray samples and 65 control microarray samples. Then, a null-hypothesis NGF model was constructed based on these permuted samples and AraONE. We repeated this computation 100 times. For each real IS, a \( t \)-test was used to compare it with the corresponding 100 ISs from the null-hypothesis. A Bonferroni correction was used to correct the \( p \)-values. A gene or interaction with a corrected \( p \)-value < 0.05 for its IS was deemed significantly important.

The classification performance of NGF was assessed through a 10-fold cross-validation test. In brief, all of the positive samples (PTI/ETI) and negative samples (control) were first randomly divided into ten subsets of the same size, respectively. The ratio of positive to negative samples in each subset was roughly the
same as that of the original (i.e., 69:65 for the PTI group and 36:65 for the ETI group).

Next, the first nine subsets were used to train a machine-learning model, and the remaining one subset was used to test the performance of the established model. This operation was repeated ten times, in which every subset was used in turn to test the model, to estimate the overall performance of the NGF. To reduce the influence of small sample size on performance stability, the 10-fold cross-validation test was repeated ten times using different dataset partitions.

**Extracting decision rules and computing interaction logic state scores**

Decision rules were extracted by traversing decision paths for PTI (ETI) of all decision trees in the NGF model and further processed using three operations to remove unreliable, inconsistent and redundant rules: (1) rules covered less than five samples were removed; (2) if there were multiple highly similar rules, only one was retained; and (3) all contradictory rules whose decision conditions were very similar but yield opposite outcomes were removed. The logic state score of a gene $g$ is defined as follows:

$$L_g = \frac{n_{up}}{N_{rule}} - 0.5 \quad \text{(8)}$$

where $n_{up}$ is the number of decision rules where $g$ is set as up-regulated. $N_{rule}$ is the total number of decision rules. The logic state score of the interaction is the following:

$$L_{12}(g_1, g_2) = \alpha \sqrt{L_{g_1}(g_1) L_{g_2}(g_2)} \quad \text{(9)}$$

where $\alpha$ is 1 if $L_{g_1}(g_1) L_{g_2}(g_2) \geq 0$; otherwise, $\alpha$ is -1.

**Detection of network modules**

MCL was used to detect network modules for the following reasons. First, it is robust to noisy biological network data (Vlasblom and Wodak, 2009) and has been successfully used in previous plant molecular network research (Tzfadia et al., 2012). Second, it has shown to be more accurate in comparison to other algorithms (Nepusz et al., 2012). Third, it is scalable to the network with thousands of nodes. The MCL algorithm (http://micans.org/mcl/) with default parameters (granularity = 2 and edge
weight =1) was used to cluster AraONE, and only network modules with a size ≥3 genes were retained.

**Identification and functional classification of significant modules**

The original GSEA algorithm (Mootha et al., 2003) was implemented with custom code to identify network modules significantly enriched with high IS genes in the PTI group (or the ETT group). Using the PTI group as an example, the NGF-based GSEA was conducted as follow. In the first step, all $N$ genes in AraONE were arranged in descending order according to their NGF ISs. In the second step, for a module with $M$ genes, a running sum was computed from the gene with the highest IS($g_1$) to that with the lowest IS($g_N$). The running sum was defined as $\sum_{i=1}^{j} s_i \ (i \leq j \leq N)$. $s_i = \sqrt{(N - M) / M}$ if $g_i$ belongs to the module; otherwise, $s_i = -\sqrt{M / (N - M)}$. In the third step, an enrichment score (ES) of the module was defined as $\max_{rs/j/N} \sum_{i=r}^{j} s_i$. To determine the statistical significance of the ESs, 100 random NGF models were built on gene expression profiles whose phenotype labels were randomly permuted. The above three computational procedures were repeated on ISs from each random model. Then, the ESs of all modules from all 100 random NGF models were pooled together to construct a background distribution of ESs. The $p$-value of each real ES was computed as the proportion of the ES in the background distribution that was equal to or larger than the real ES (Mootha et al., 2003). Modules with an empirical $p$-value < 0.05 were considered significant.

We performed a functional annotation of the significant modules using BiNGO and classified them into five functional categories: (1) signal transduction, (2) transcriptional regulation, (3) hormone, (4) defense response, and (5) growth or development. Because there could be more than one enriched functional term within a module, we only considered the most significant term (i.e., the GO term with the lowest $p$-value). The modules that were annotated with terms containing the words “signal transduction” or “phosphorylation” were partitioned into category (1); those
that were annotated with terms including “regulation of transcription” or “RNA processing” were partitioned into category (2); those that were annotated with hormone-related terms such as “response to auxin” or “ethylene-mediated signaling pathway”, or including known master hormone regulators, such as NPR1 (Dong, 2004), were grouped into category (3); those that were annotated with terms including “defense response”, “hypersensitive response” or “response to bacterium” or important defense-related process, such as “indole glucosinolate biosynthetic process” (Clay et al., 2009), were partitioned into category (4); and those that were annotated with terms including “photosynthesis”, “growth”, “development”, “cell cycle”, or “organ morphogenesis” were partitioned into category (5). The modules that were annotated with other functions were removed. A complete list of the significant modules is provided in Table S10.

**Calculation of closeness centrality**

As an index to measure the node’s closeness to the center of the network, *CC* is defined as follows:

\[
CC(x) = \frac{N - 1}{\sum_{y \in U, y \neq x} d(x, y)}
\]  

where *N* is the gene number in the network, *U* is the set of all genes in the network, and *d(x,y)* is the shortest network distance between genes *x* and *y*.

**Network visualization**

Cytoscape (Saito et al., 2012) and its plugins were used to visualize the network. Specifically, Cerebral plugin (http://www.pathogenomics.ca/cerebral/) was used to visualize the shared subnetwork in pathway style. The subcellular location information of gene production was downloaded from SUBA3 (http://suba.plantenergy.uwa.edu.au/). The GO enrichment results from BiNGO were imported into the EnrichmentMap plugin (http://www.baderlab.org/Software/EnrichmentMap) to create an Enrichment Map of the shared network.
ACKNOWLEDGEMENTS

We thank the editor and anonymous referees whose constructive comments were very helpful in improving the quality of this work. We are also grateful to Drs. Wenxian Sun and Changfu Tian at China Agricultural University, and Dr. Fei He at Brookhaven National Laboratory (USA) for their critical reading and helpful comments.

LITERATURE CITED

Dong X (2004) NPR1, all things considered. Current Opinion in Plant Biology 7: 547-552


Plant Journal 50: 347-363
Mechanisms underlying robustness and tunability in a plant immune signaling network. Cell Host & Microbe 15: 84-94
(2003) PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately

T, Nishimura MT, Pevzner SJ, Donovan SE, Ghamsari L, Santhanam B, Romero V,
Poulin MM, Gebreab F, Gutierrez BJ, Tam S, Monachello D, Boxem M, Harbort CJ,
McDonald N, Gai L, Chen H, He Y, Consortium EUE, Vandenhaute J, Roth FP, Hill DE,
effectors converge onto hubs in a plant immune system network. Science 333: 596-601

view on networking by hormones in Arabidopsis immunity reveals multiple crosstalk for
cytokinin. The Plant Cell

Nepusz T, Yu H, Paczynski A (2012) Detecting overlapping protein complexes in protein-

Pajerowska-Mukhtar Karolina M, Wang W, Tada Y, Oka N, Tucker Chandra L, Fonseca Jose P,
Dong X (2012) The HSF-like Transcription Factor TBF1 Is a Major Molecular Switch for
Plant Growth-to-Defense Transition. Current Biology 22: 103-112

Pandey SP, Somssich IE (2009) The role of WRKY transcription factors in plant immunity. Plant
Physiology 150: 1648-1655


modeling reveals prevalent negative regulatory relationships between signaling sectors in
Arabidopsis immune signaling. PLoS Pathog 6: e1001011

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D,
37: 501-506

Phosphorylation-dependent differential regulation of plant growth, cell death, and innate
immunity by the regulatory receptor-like kinase BAK1. PLoS Genet 7: e1002046

Shaik R, Ramakrishna W (2013) Genes and co-expression modules common to drought and bacterial
stress responses in Arabidopsis and rice. PLoS ONE 8: e77261

Shaik R, Ramakrishna W (2014) Machine learning approaches distinguish multiple stress conditions
using stress-responsive genes and identify candidate genes for broad resistance in rice. Plant
Physiology 164: 481-495

cells. Nat Rev Immunol 12: 89-100

108: 451-457

Quantitative nature of Arabidopsis responses during compatible and incompatible interactions
with the bacterial pathogen Pseudomonas syringae. The Plant Cell 15: 317-330


Truman W, de Zabala MT, Grant M (2006) Type III effectors orchestrate a complex interplay between transcriptional networks to modify basal defence responses during pathogenesis and resistance. The Plant Journal 46: 14-33


Yoo SY, Kim Y, Kim SY, Lee JS, Ahn JH (2007) Control of flowering time and cold response by a
NAC-domain protein in *Arabidopsis*. PLoS ONE 2: e642


**Figure Legends**

**Figure 1.** Methodological overview of the integrative network analyses. The analyses consisted of five steps (from left to right). First, multiple genomic data were collected to assemble an integrated gene network with ~11,000 nodes and ~64,000 edges. Second, an NGF algorithm was used to combine gene expression data with the gene network to rank the importance of the network components (genes and interactions) in PTI and ETI. Third, a shared subnetwork between PTI and ETI was identified and validated using independent data. The global network organizations of PTI and ETI were compared using modular network models. An interactive network web browser was developed for the convenience of the scientific community.

**Figure 2.** Comparison of NGF with differential expression analysis and characterization of the topological properties of significantly important gene interactions. (A) Change (absolute value) of gene expression in the PTI group. (B) Change (absolute value) of gene expression in the ETI group. (C) NGF important scores (ISs) of genes in the PTI group. (D) NGF ISs of genes in the ETI group. The top 30 DEGs (green) and the top 30 genes with the highest NGF ISs (purple) are shown. The genes included in the top 30 genes of both the DEGs and the NGF results are indicated in orange. Overall, the top 30 genes ranked by NGF tend to be differentially expressed, but some immune-related genes even with moderate expression alterations, such as *CPK1* and *NPR1* (red), are also discriminated from the noisy background as a top 30 gene by NGF. The genes are ordered along the X-axis according to their AGI numbers. In both the PTI (E) and ETI (F) groups, the significantly important gene interactions identified by NGF tend to link different network modules. This phenomenon is more obvious when the IS cutoff is increased.
**Figure 3.** The shared subnetwork between PTI and ETI. (A) The subnetwork is arranged in traditional pathway style using the Cerebral plugin according to the confirmed subcellular location information of gene products from SUBA3. Forty-nine putative effector targets (red) are enriched in this subnetwork ($p$-value=6.44×10^{-15}, hypergeometric test), and 19 WRKY family TFs (green triangles) are included. (B) Enrichment map of the shared subnetwork. Each node represents a gene set corresponding to a GO term enriched in the shared subnetwork, and each weighted edge represents the overlap between two gene sets. The node size is proportional to the number of genes that belong to this GO term. The width of the edge is proportional to the size of the overlapping gene sets.

**Figure 4.** Modular network models of PTI and ETI. Each node represents a network module identified by NGF-based GSEA. The circle nodes are modules common to the PTI and ETI models, and the triangle nodes are modules that are specific to the corresponding models. The width of the module interaction is proportional to the average interaction IS as assigned by NGF. The modules and interactions are colored according to the average logic state score. The modules are presented according to five functional categories. The detailed gene composition of the modules and their interactions are available at http://systbio.cau.edu.cn/pinet/home.php

**Figure 5.** Organization of defense response modules in PTI and ETI. (A) The defense response modules are organized into a compact clique with signal transduction modules in PTI. Positive combinations of logic states are prevalent among the modules. (B) Defense response modules are relatively independent of each other, and different parts of network models are mainly linked by a protein phosphorylation module. The color scheme and the meaning of each shape are the same as in Figure 4.
**Tables**

Table 1. Prediction performance of NGF$^a$

<table>
<thead>
<tr>
<th></th>
<th>Acc</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTI vs. control</td>
<td>0.859 ± 0.012</td>
<td>0.942 ± 0.008</td>
</tr>
<tr>
<td>ETI vs. control</td>
<td>0.857 ± 0.014</td>
<td>0.934 ± 0.011</td>
</tr>
</tbody>
</table>

$^a$The number of trees is 100.
Table 2. Top 30 genes in PTI vs. control.

<table>
<thead>
<tr>
<th>AGI number</th>
<th>Name</th>
<th>Function in Plant Immunity*</th>
<th>NGF IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g88520</td>
<td>B-BOX DOMAIN PROTEIN 14, BBX14</td>
<td>SA-mediated signaling pathway</td>
<td>0.660</td>
</tr>
<tr>
<td>At2g02450</td>
<td>LONG VEGETATIVE PHASE1, LOV1</td>
<td>unknown</td>
<td>0.573</td>
</tr>
<tr>
<td>At4g23810</td>
<td>WRKY DNA-BINDING PROTEIN 53, WRKY53</td>
<td>coordinates with WRKY46 in basal resistance</td>
<td>0.453</td>
</tr>
<tr>
<td>At4g18880</td>
<td>HEAT SHOCK TRANSCRIPTION FACTORA4A, HSFA4A</td>
<td>response to chitin; substrate of the MPK3/MPK6 signaling</td>
<td>0.382</td>
</tr>
<tr>
<td>At4g01090</td>
<td>unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At1g72610</td>
<td>GERMIN-LIKE PROTEIN GLP1</td>
<td>unknown</td>
<td>0.379</td>
</tr>
<tr>
<td>At2g04570</td>
<td>WRKY DNA-BINDING PROTEIN 17, WRKY17</td>
<td>negative regulator of basal resistance</td>
<td>0.361</td>
</tr>
<tr>
<td>At5g18150</td>
<td>unknown</td>
<td></td>
<td>0.346</td>
</tr>
<tr>
<td>At1g14370</td>
<td>PBS1-LIKE 2 PROTEIN KINASE 2A, PBL2</td>
<td>cleaved by AvrPphB; pbl2 mutant showed significantly reduced callose deposition in flg22 and elf18 treatment</td>
<td>0.340</td>
</tr>
<tr>
<td>At1g69770</td>
<td>CHROMOMETHYLASE 3, CMT3</td>
<td>involved in defense response-related DNA methylation</td>
<td>0.303</td>
</tr>
<tr>
<td>At3g26085</td>
<td>unknown</td>
<td></td>
<td>0.276</td>
</tr>
<tr>
<td>At3g46780</td>
<td>PLASTID TRANSCRIPTIONALLY ACTIVE 16, PTAC16</td>
<td></td>
<td>0.275</td>
</tr>
<tr>
<td>At5g04870</td>
<td>CALCIUM DEPENDENT PROTEIN KINASE 1, CPK1</td>
<td>control onset of cell death triggered by RPS2 and RPM1</td>
<td>0.271</td>
</tr>
<tr>
<td>At2g38470</td>
<td>WRKY DNA-BINDING PROTEIN 33, WRKY33</td>
<td>key transcriptional regulator of and hormonal and metabolic responses to necrotroph</td>
<td>0.268</td>
</tr>
<tr>
<td>At1g34310</td>
<td>AUXIN RESPONSE FACTOR 12, ARF12</td>
<td>unknown</td>
<td>0.259</td>
</tr>
<tr>
<td>At3g17700</td>
<td>CYCLIC NUCLEOTIDE GATED CHANNEL 20, CNGC20</td>
<td>regulation of plant-type hypersensitive response</td>
<td>0.257</td>
</tr>
<tr>
<td>At2g29100</td>
<td>GLUTAMATE RECEPTOR 2.9, GLR2.9</td>
<td>unknown</td>
<td>0.255</td>
</tr>
<tr>
<td>At5g05190</td>
<td>unknown</td>
<td></td>
<td>0.255</td>
</tr>
<tr>
<td>At4g02410</td>
<td>LECTIN-LIKE PROTEIN KINASE 1, LPK1</td>
<td>confers Arabidopsis pathogen resistance when overexpressed</td>
<td>0.250</td>
</tr>
<tr>
<td>At2g33500</td>
<td>B-BOX DOMAIN PROTEIN 12, BBX12</td>
<td>unknown</td>
<td>0.247</td>
</tr>
<tr>
<td>At5g25930</td>
<td>unknown</td>
<td></td>
<td>0.245</td>
</tr>
<tr>
<td>At4g09570</td>
<td>CALCIUM DEPENDENT PROTEIN KINASE 4, CPK4</td>
<td>directly phosphorlates WRKY8/28/48</td>
<td>0.244</td>
</tr>
<tr>
<td>At4g14540</td>
<td>NUCLEAR FACTOR Y, SUBUNIT B3, NF-YB3</td>
<td>unknown</td>
<td>0.243</td>
</tr>
<tr>
<td>At2g30250</td>
<td>WRKY DNA-BINDING PROTEIN 25, WRKY25</td>
<td>negative regulator of SA-mediated defense responses</td>
<td>0.238</td>
</tr>
<tr>
<td>At3g59060</td>
<td>PHYTOCHROME INTERACTING FACTOR 3-LIKE 6, PIL6</td>
<td>unknown</td>
<td>0.235</td>
</tr>
<tr>
<td>At3g07040</td>
<td>RESISTANCE TO P. SYRINGAE PV. MACULICOLA 1, RPM1</td>
<td>R gene</td>
<td>0.233</td>
</tr>
<tr>
<td>At1g35420</td>
<td>unknown</td>
<td></td>
<td>0.229</td>
</tr>
<tr>
<td>At5g11060</td>
<td>KNOTTED1-LIKE HOMEOBOX GENE 4, KNAT4</td>
<td>unknown</td>
<td>0.228</td>
</tr>
<tr>
<td>At3g55250</td>
<td>PIGMENT DEFECTIVE 329, PDE329</td>
<td>unknown</td>
<td>0.224</td>
</tr>
<tr>
<td>At1g75460</td>
<td>unknown</td>
<td></td>
<td>0.222</td>
</tr>
</tbody>
</table>

*Complete functional annotation of each gene and the corresponding references are provided in Table S6.
Table 3. Top 30 genes in ETI vs. control.

<table>
<thead>
<tr>
<th>AGI number</th>
<th>Name</th>
<th>Function in Plant Immunitya</th>
<th>NGF IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>At3g44720</td>
<td>AROGENATE DEHYDRATASE 4, ADT4</td>
<td>regulation of the plant-type hypersensitive response</td>
<td>0.275</td>
</tr>
<tr>
<td>At3g57150</td>
<td>HOMOLOGUE OF NAP57, NAP57</td>
<td>unknown</td>
<td>0.270</td>
</tr>
<tr>
<td>At5g46760</td>
<td>MYC3</td>
<td>mediates the JA response</td>
<td>0.215</td>
</tr>
<tr>
<td>At1g63100</td>
<td></td>
<td>unknown</td>
<td>0.212</td>
</tr>
<tr>
<td>At2g02450</td>
<td>LONG VEGETATIVE PHASE 1, LOV1</td>
<td>unknown</td>
<td>0.212</td>
</tr>
<tr>
<td>At2g13650</td>
<td>GOLGI NUCLEOTIDE SUGAR TRANSPORTER 1, GONST1</td>
<td>gnost1 plants have a dwarfed phenotype and a constitutive hypersensitive response with elevated salicylic acid levels</td>
<td>0.199</td>
</tr>
<tr>
<td>At1g59590</td>
<td>ZCF37</td>
<td>unknown</td>
<td>0.197</td>
</tr>
<tr>
<td>At2g30250</td>
<td>WRKY DNA-BINDING PROTEIN 25, WRKY25</td>
<td>negative regulator of SA-mediated defense responses</td>
<td>0.174</td>
</tr>
<tr>
<td>At1g14370</td>
<td>PBSL1-LIKE PROTEIN KINASE 2A, PBL2</td>
<td>cleaved by AvrPphB; pbl2 mutant exhibits significantly reduced callose deposition upon flg22 and elf18 treatment</td>
<td>0.173</td>
</tr>
<tr>
<td>At5g52010</td>
<td></td>
<td>unknown</td>
<td>0.173</td>
</tr>
<tr>
<td>At3g22780</td>
<td>CHINESE FOR &quot;UGLY&quot;, TSO1</td>
<td>unknown</td>
<td>0.169</td>
</tr>
<tr>
<td>At3g59220</td>
<td>PRN1</td>
<td>unknown</td>
<td>0.165</td>
</tr>
<tr>
<td>At4g18880</td>
<td>HEAT SHOCK TRANSCRIPTION FACTOR A4A, HSF A4A</td>
<td>response to chitin; substrate of MPK3/MPK6 signaling</td>
<td>0.160</td>
</tr>
<tr>
<td>At3g56400</td>
<td>WRKY DNA-BINDING PROTEIN 70, WRKY70</td>
<td>convergence between SA and JA signaling</td>
<td>0.158</td>
</tr>
<tr>
<td>At1g67690</td>
<td></td>
<td>unknown</td>
<td>0.155</td>
</tr>
<tr>
<td>At2g38470</td>
<td>WRKY DNA-BINDING PROTEIN 33, WRKY33</td>
<td>key transcriptional regulator of hormonal and metabolic responses to necrotroph</td>
<td>0.149</td>
</tr>
<tr>
<td>At1g31190</td>
<td>MYO-INOSITOL MONOPHOSPHATASE LIKE 1, IMPL1</td>
<td>unknown</td>
<td>0.149</td>
</tr>
<tr>
<td>At5g20950</td>
<td></td>
<td>unknown</td>
<td>0.147</td>
</tr>
<tr>
<td>At4g19030</td>
<td>NOD26-LIKE MAJOR INTRINSIC PROTEIN 1, NLM1</td>
<td>regulation of the defense response</td>
<td>0.146</td>
</tr>
<tr>
<td>At1g71260</td>
<td>WHIRY 2, WHY2</td>
<td>defense response</td>
<td>0.146</td>
</tr>
<tr>
<td>At1g64280</td>
<td>NON-EXPRESSER OF PR GENES 1, NPR1</td>
<td>master regulator of SA-mediated systemic acquired resistance</td>
<td>0.145</td>
</tr>
<tr>
<td>At3g16870</td>
<td>GATA TRANSCRIPTION FACTOR 17, GATA17</td>
<td>unknown</td>
<td>0.139</td>
</tr>
<tr>
<td>At1g29120</td>
<td></td>
<td>unknown</td>
<td>0.138</td>
</tr>
<tr>
<td>At1g36310</td>
<td></td>
<td>unknown</td>
<td>0.137</td>
</tr>
<tr>
<td>At1g17440</td>
<td>CYTOKININ-HYPERSENSITIVE 1, ENHANCED ETHYLENE RESPONSE 4, CKH1/EER4</td>
<td>unknown</td>
<td>0.130</td>
</tr>
<tr>
<td>At3g48090</td>
<td>ENHANCED DISEASE SUSCEPTIBILITY 1, EDS1</td>
<td>indispensable for immunity mediated by TIR-NB-LRR receptors; acts redundantly with SA to regulate R gene-mediated signaling</td>
<td>0.130</td>
</tr>
<tr>
<td>At4g29810</td>
<td>MAP KINASE KINASE 1, MKK2</td>
<td>negative regulation of the defense response</td>
<td>0.130</td>
</tr>
<tr>
<td>At1g28480</td>
<td>GRX480/ROXY19</td>
<td>negative regulation of the defense response</td>
<td>0.129</td>
</tr>
<tr>
<td>At3g02910</td>
<td></td>
<td>unknown</td>
<td>0.127</td>
</tr>
<tr>
<td>At3g10640</td>
<td>VPS60.1</td>
<td>unknown</td>
<td>0.123</td>
</tr>
</tbody>
</table>

*Complete functional annotation of each gene and the corresponding references are provided in Table S6
Supplemental Data

Figure S1. Assembly of the integrative gene network (AraONE).

Figure S2. The cumulative distribution of reliability score in the PPI data.

Figure S3. Comparison of the GO term annotation similarities of AraONE/PPI and the corresponding random network.

Figure S4. The average GO term annotation similarity of retained PPIs as a function of PPI reliability score.

Figure S5. Venn diagram of the overlap among the five datasets.

Figure S6. Reproducibility of NGF.

Figure S7. Differential expression of NPR1, CPK1 and their neighbors.

Figure S8. Force-directed layout of the shared subnetwork.

Figure S9. Biological processes enriched in the shared subnetwork after removing the interactions inferred from gene co-expression data.

Figure S10. Functional annotations of module M63.

Figure S11. Sub-modules derived from M63.

Figure S12. Closeness centrality of M63.

Table S1. Source of the chromatin modification data.

Table S2. AraONE interactions annotated with their data sources.

Table S3. Composition of the gene expression profiles.

Table S4. Performance of NGF on two independent datasets.

Table S5. All significantly important genes and interactions identified by NGF.

Table S6. Extended Table 2 and Table 3 with functional annotation and reference information for each gene.
Table S7. The top 30 DEGs and their functional annotation.

Table S8. Gene members of the 633 network modules.

Table S9. List of 49 putative effector targets within the shared subnetwork.

Table S10. Significant network modules identified by NGF-based GSEA and their functional annotations.

Table S11. Functional annotation of the network modules of the two modular models.

Data S1. The shared subnetwork in Cytoscape (.cys) format. Cytoscape 2.8.2 installed with the Cerebral plugin is recommended to explore this dataset.
Figure 1. Methodological overview of the integrative network analyses. The analyses consisted of five steps (from left to right). First, multiple genomic data were collected to assemble an integrated gene network with ~11,000 nodes and ~64,000 edges. Second, an NGF algorithm was used to combine gene expression data with the gene network to rank the importance of the network components (genes and interactions) in PTI and ETI. Third, a shared subnetwork between PTI and ETI was identified and validated using independent data. The global network organizations of PTI and ETI were compared using modular network models. An interactive network web browser was developed for the convenience of the scientific community.
Figure 2. Comparison of NGF with differential expression analysis and characterization of the topological properties of significantly important gene interactions. (A) Change (absolute value) of gene expression in the PTI group. (B) Change (absolute value) of gene expression in the ETI group. (C) NGF important scores (ISs) of genes in the PTI group. (D) NGF ISs of genes in the ETI group. The top 30 DEGs (green) and the top 30 genes with the highest NGF ISs (purple) are shown. The genes included in the top 30 genes of both the DEGs and the NGF results are indicated in orange. Overall, the top 30 genes ranked by NGF tend to be differentially expressed, but some immune-related genes even with moderate expression alterations, such as *CPK1* and *NPR1* (red), are also discriminated from the noisy background as a top 30 gene by NGF. The genes are ordered along the X-axis according to their AGI numbers. In both the PTI (E) and ETI (F) groups, the significantly important gene interactions identified by NGF tend to link different network modules. This phenomenon is more obvious when the IS cutoff is increased.
Figure 3. The shared subnetwork between PTI and ETI. (A) The subnetwork is arranged in
traditional pathway style using the Cerebral plugin according to the confirmed subcellular location information of gene products from SUBA3. Forty-nine putative effector targets (red) are enriched in this subnetwork ($p$-value=$6.44 \times 10^{-15}$, hypergeometric test), and 19 WRKY family TFs (green triangles) are included. (B) Enrichment map of the shared subnetwork. Each node represents a gene set corresponding to a GO term enriched in the shared subnetwork, and each weighted edge represents the overlap between two gene sets. The node size is proportional to the number of genes that belong to this GO term. The width of the edge is proportional to the size of the overlapping gene sets.
Figure 4. Modular network models of PTI and ETI. Each node represents a network module identified by NGF-based GSEA. The circle nodes are modules common to the PTI
and ETI models, and the triangle nodes are modules that are specific to the corresponding models. The width of the module interaction is proportional to the average interaction IS as assigned by NGF. The modules and interactions are colored according to the average logic state score. The modules are presented according to five functional categories. The detailed gene composition of the modules and their interactions are available at http://systbio.cau.edu.cn/pinet/home.php.
Figure 5. Organization of defense response modules in PTI and ETI. (A) The defense response modules are organized into a compact clique with signal transduction modules in PTI. Positive combinations of logic states are prevalent among the modules. (B) Defense response modules are relatively independent of each other, and different parts of network models are mainly linked by a protein phosphorylation module. The color scheme and the meaning of each shape are the same as in Figure 4.