IDENTIFICATION OF TRANSCRIPTOME DYNAMIC PATTERNS AND THEIR COGNATE CIS-ELEMENTS IN DEFENSE GENES IN ARABIDOPSIS

by

Oluwadamilare I. Afolabi, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Biology
December 2020

Committee Members:

Hong-Gu Kang, Chair

Nihal Dharmasiri

Sunethra Dharmasiri

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

DEDICATION

This thesis is dedicated to my mother and the Almighty God, who have been my source of inspiration and strength.

ACKNOWLEDGEMENTS

I would like to express my deep and sincere gratitude to my research supervisor, Dr. Hong-Gu Kang, for giving me the opportunity to research in his lab and providing guidance throughout my research. I would also like to thank other committee members, Dr. Nihal Dharmasiri and Dr. Sunethra Dharmasiri, for their valuable time and critical feedback throughout this research.

I am extending my heartfelt gratitude to Dr Sung-Il Kim, Dr Yogendra Bordiya, and other lab members, for their assistance and feedback during my research. I would not have made it but for the department of Biology's funding for my program for which I am very grateful.

Lastly, I am extremely grateful to my mother and other family members for their many intercessions, love, and support towards my education and career.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

Abbreviation	Description
AME	Analysis of motif enrichment
BAM	Binary Alignment Map
bHLH	Basic helix-loop-helix
bp	Base pair
bZIP	Basic leucine zipper
CDS	Coding sequence
CFU	Colony-forming unit
Col-0	Columbia ecotype
DAP-Seq	DNA Affinity Purification and sequencing
DCL	Dicer-like protein
DEG	Differentially expressed gene
DNA	Deoxyribonucleic acid
Dof	DNA binding with one finger
EF-Tu	Thermo-unstable elongation factor
ETI	Effector-triggered immunity
ETS	Effector-triggered susceptibility
FDR	False discovery rate
FLS2	Flagellin-sensitive 2
FPKM	Fragments Per Kilobase per Million

GO	Gene Ontology
hpi	Hours post-infection
HR	Hypersensitive response
JAZ	Jasmonate-zim
LRR	Leucine rich repeat
MAP	Mitogen-activated protein
MEME	Multiple Expression motifs for Motif Elicitation
$MgCl_2$	Magnesium chloride
NB	Nucleotide binding
NGS	Next-generation sequencing
PAMP	Pathogen-associated molecular pattern
PCA	Principal component analysis
PIF	Phytochrome interacting factors
PR	Pathogenesis-related
PRR	Pattern recognition receptors
PTI	PAMP-triggered immunity
QC	Quality control
R	Resistance
R	The R Project for Statistical Computing
RASL-Seq	RNA-mediated oligonucleotide annealing, selection, and ligation with next-generation sequencing

RIN4	RPM1 interacting protein 4
RIPK	RPM1-induced protein kinase
RLK	Receptor-like kinase
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
ROC1	Rotamase CYP 1
RPM1	Resistance to *P. syringae pv. maculicola* 1
RPS2	Resistant to P. *syringae* 2
RRS1	Resistant to ralstonia solanacearum 1
SA	Salicylic acid
SAM	Sequence Alignment Map
SOM	Self-organizing map
TALE	Transcription activator-like effector
TF	Transcription factors
TMM	Trimmed Mean of M-values
TSS	Transcriptional start site
ZOOPS	Zero or One Occurrence per Sequence
3D	Three-dimensional

ABSTRACT

Successful defense responses depend on the timely detection of pathogens and rapid transcriptional changes of defense genes. These transcriptional changes are mainly controlled by transcription factors binding to the regulatory regions of defense genes. To track these transcriptional changes and group these into a few related patterns, I analyzed the genome-wide expression profiles of *Arabidopsis thaliana* plants responding to *Pseudomonas syringae* infection at multiple times points. There are two distinct groups of defense genes rapidly upregulated or downregulated, reacting to *Pseudomonas* infection. I clustered these highly dynamical genes by their expression patterns into nine groups. I analyzed the gene ontology of each cluster and found that up-regulated clusters were mostly represented by traditional defense genes and down-regulated ones included a significant number of photosynthesis-related genes. Assessment of the promoter sequences also found that upregulated and downregulated clusters have distinct groups of *cis*-elements and their potential binding factors. In sum, my thesis research identified nine clusters with separable expression patterns and suggested a few candidate master regulators in transcriptional reprogramming associated with biotic stress in Arabidopsis.

I. INTRODUCTION

Molecular Mechanism of Arabidopsis-*Pseudomonas syringae* Interaction

The ability of plants to defend themselves against pathogens is truly remarkable, given that they are incessantly exposed to pathogens and lack dedicated immune cells and adaptive somatic variations. Instead, plants depend upon a combination of defense mechanisms to defend themselves against a wide range of potential pathogens in an autonomous fashion. In addition to physical and chemical defenses, plant Pattern Recognition Receptors (PRRs) recognize conserved pathogen-associated molecular patterns (PAMPs) to induce PAMP-triggered immunity (PTI) (Jones & Dangl, 2006; Chisholm *et al.*, 2006; Sun & Zhang, 2020). Bacterial flagellin is one of the most studied PAMPs in plants. In Arabidopsis, the receptor-like kinase FLS2 elicits a defense response when it recognizes a conserved, 22-amino acid epitope (flg22) of bacterial flagellin (Zipfel *et al.*, 2004; Helft *et al*., 2016). Like flagellin, many other bacterial PAMPs are recognized by plant PRRs. These include bacterial cell wall and cell membrane components, bacterial EF-Tu, and bacterial nucleic acids (Newton *et al*., 2010; Lee *et al.*, 2016). Responses associated with PTI include increased cytosolic Ca^{2+} levels and activation of MAP kinase cascades (Blume *et al*., 2000; Ranf *et al*., 2011).

Pathogens dampen or evade these PTI responses by secreting effector molecules or virulence factors into plants (Figure 1). The highly virulent *Pseudomonas syringae* pv. *tomato* DC3000 (*Pst*), which causes disease in tomato, Arabidopsis, and *Nicotiana benthamiana*, possesses over 30 effectors that are delivered by the type III secretion system (Chang *et al.*, 2005; Schechter *et al.*, 2006; Cunnac *et al.*, 2011). R proteins specifically recognize these effectors to elicit a more rapid and robust immune response

known as effector-triggered immunity (ETI) (Jones & Dangl, 2006; Stuart *et al*., 2013). In contrast to PTI, ETI often results in the hypersensitive response (HR), a programmed cell death around the area of infection, which diminishes the nutrient supply to the pathogen and restrict its mobility, preventing further infection to other parts of the plant (Jones & Dangl, 2006; Kang *et al*., 2008). Additionally, HR is associated with an increase in the defense signaling molecule, salicylic acid (SA), and defense gene induction, including several pathogenesis-related (*PR*) genes (Nazar *et al*., 2017; Hadwiger & Tanka, 2017). Contrary to early expectations that R proteins would be receptors for effectors, many R proteins detect effectors indirectly, by sensing changes in other host proteins that are targeted by pathogen-derived effectors (Van Der Biezen & Jones, 1998; DeYoung & Innes, 2006).

In Arabidopsis, RIN4 interacts with the disease resistance proteins RPS2 and RPM1. The phosphorylation of RIN4 by RPM1-induced protein kinase (RIPK) kinase is mediated by the *P. syringae* Type-III effectors AvrB and AvrRpm1 (Liu *et al*., 2011; Mackey *et al*., 2002). The phosphorylation of RIN4 at threonine residue (T166) inhibits RIN4 interaction with ROC1, resulting in conformational changes of RIN4, which initiates RPM1-mediated immune signaling (Li *et al*., 2014). Alternatively, direct cleavage of RIN4 by another *P. syringae* effector AvrRpt2 inhibits RPM1-mediated immune signaling (Afzal *et al*., 2011). RIN4 binding to RPS2 keeping RPS2 inactivated, but AvrRpt2-mediated cleavage releases the inhibition of RPS2, triggering immune signaling (Day *et al*., 2005; Kim *et al*., 2009).

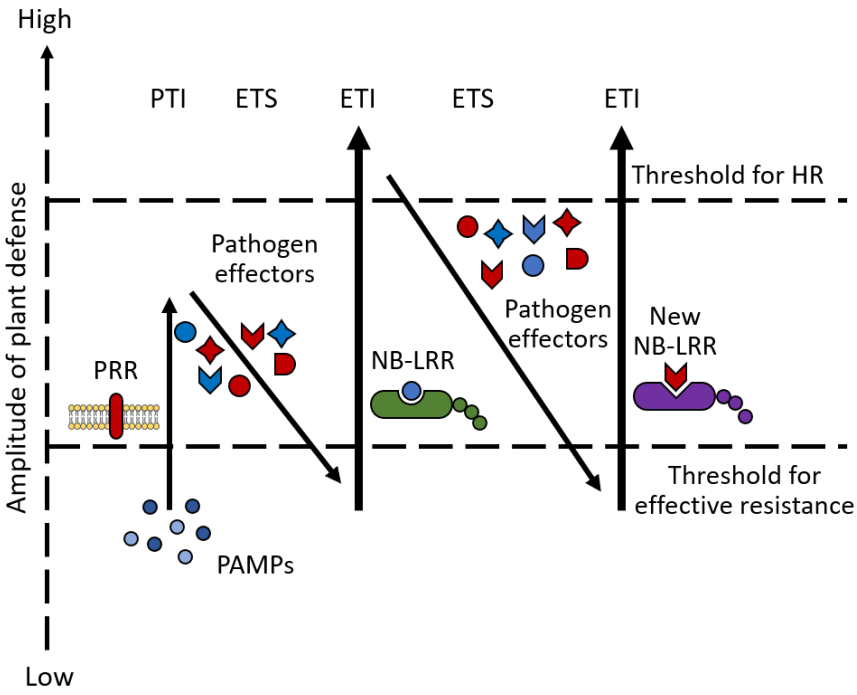


Figure 1. The Zig-Zag Model as Described by Jones & Dangl (2006). This schematic diagram of plant innate immunity illustrates sequential plant-pathogen interactions. The process begins with the recognition of the pathogen-associated molecular patterns (PAMPs) by their cognate PRRs generating PTI (PAMP triggered immunity). Pathogen effectors can prevent PAMP recognition, resulting in effector-triggered susceptibility (ETS). These effectors become avirulence factors if they are detectable by plant NB-LRR protein leading to effector-triggered immunity (ETI). ETS and ETI alternate perpetually as pathogens loose old effectors and/or acquire new effectors, and plant develops new NB-LRR alleles for the new effectors, respectively.

There are many other mechanisms by which plants recognize the presence of effectors. For instance, the C-terminal WRKY domain of Arabidopsis RRS1 interacts with *P. syringae* Type-III effector AvrRps4, activating an RPS4-dependent immune response (Sohn *et al*., 2012). However, the purpose of RRS1 WRKY domain binding to W-box DNA sequence is not known (Sarris *et al*., 2015). Also, executor genes are R genes that are transcriptionally activated by transcription activator-like effectors (TALEs) produced by different *Xanthomonas* species. TALEs bind to host regulatory DNA sequences, promoting host transcription of key susceptibility factors. Executor gene

promoters mimic the promoter regions of these susceptibility factors, leading to the initiation of defense responses (Malik & Van der Hoorn, 2016). So far, executor genes have only been identified in rice and pepper. However, understanding the sequence specificity of TALEs has enabled researchers to develop synthetic executor genes mediating immunity against *Xanthomonas* species and other TALE expressing pathogens (Ji *et al*., 2016; Ji *et al*., 2018).

Dynamics of Transcription Under Biotic Stress and its Analysis via NGS (Next-Generation Sequencing)

Unlike vertebrate immune systems, individual plant cells must respond to challenges from pathogens and defense signals from neighboring cells. Mounting a successful defense response to pathogen infection depends heavily on the timely and accurate detection of the pathogen and the subsequent induction of defense response genes (Hoang *et al*., 2017; Nath *et al*., 2019). The speed and specificity of the resulting global transcriptional changes indicate a significant role of transcription factors (TFs) in controlling these changes. Recent studies have considerably enhanced our understanding of the genome-wide transcriptional landscape controlling plant defense responses and revealed the roles of pathway-specific TFs in the regulation of defense gene expression. The coordinated action of gene-specific TFs and the general transcriptional machinery contributes to the defense gene response (Li *et al*., 2016). Also, defense signaling components are often transcriptionally upregulated upon MAMP or effector perception to ensure robust and sustained defense response. Thus, understanding the transcriptional dynamics associated with defense responses under biotic stress enables researchers to engineer more effective approaches to control diseases in plants caused by pathogens. To

this end, the genome-wide expression profiles of different plants responding to different pathogens have been analyzed. Zhang *et al.* (2018), investigated the gene expression patterns of two susceptible poplar subgenera at the initial infection (6 hpi), biotrophic (36 hpi), and necrotrophic (96 hpi) phases of *Marssonina brunnea* infection and by so doing showed that transcriptome dynamics correlates with disease progression. Guan *et al.* (2020), in a dual RNA-Seq analysis of ginseng roots responding to *Ilyonectria robusta* infection at 36, 72, and 144 hpi, showed that gene expression patterns were highly related to physiological conditions and that defense response persisted for at least 144 hpi. Transcriptome data for Arabidopsis responding to pathogen infection, though available for many pathogens, is often inadequate. Zhu *et al*., (2013) captured Arabidopsis transcriptome dynamics following fungus *Fusarium oxysporum* infection at 1 day-post-inoculation (1 dpi) and 6 dpi. They found that most up-regulated genes at the early stages of infection mostly peaked at 6 dpi. Although Lewis *et al.* (2015) examined the high-resolution time course of genome-wide expression changes in Arabidopsis post-infection with *Pst*, it was limited to virulent and nonpathogenic strains, failing to capture expression changes during ETI. Maleck *et al.* (2001) monitored gene-expression changes in Arabidopsis under different SAR-inducing or -repressing conditions but their study. However, their study was limited to only 30% of annotated *A. thaliana* genes. Mine *et al.* (2018) examined dynamics using time-series RNA-sequencing data of Arabidopsis upon challenge with virulent and ETI-triggering avirulent strains. They found that resistant mutants and wildtype plants achieved the highest defense gene induction at 6 hpi and 24 hpi in response to avirulent and virulent *Pst* infections, respectively.

Clustering Analysis of Transcription Dynamics

Clustering analysis is a machine learning strategy whereby data observations are partitioned into groups, so that members of each group/cluster are as similar to each other as possible but as different from members of other clusters as possible. There are various clustering strategies, all of which work differently, usually yield different results, and are optimal for specific applications. Clustering algorithms can be classified into broader groups: Hierarchical clustering, Partitioning Relocation Clustering, Density-Based Partitioning, Grid-Based Methods, Co-Occurrence of Categorical Data, Constraint-Based Clustering, Relation to Supervised Learning, Gradient Descent and Artificial Neural Networks, Evolutionary Methods (Berkhin, 2006). By grouping genes (observations) with similar expression profiles across samples, cluster analyses can provide insights into gene functions and networks. However, there have only been a few published computational assessments of RNA-Seq data cluster analyses, considering the huge amount of RNA-Seq data that has been generated in recent times.

Self-Organizing Maps (SOM) is an artificial neural network that has been used for clustering. SOM does not rely on distributional assumptions, is robust to noise, handles very large datasets with ease, allows assessment of all relationships simultaneously, and is applicable to data from a wide range of sources. Different implementations of SOM have been used for both clustering and visualization of the patterns of DEGs. Cenik *et al.* (2015) trained a SOM map using gene expression (RNA-Seq) and protein level (spectrometry) data and found that each neuron within the SOM contains genes that share a similar pattern of expression and protein levels. Zhang *et al.* (2019) by functional annotation of SOM-clustered DEGs, revealed an important role for NAC7 in regulating

genes in photosynthesis, chlorophyll degradation, and protein turnover pathways that each contribute to the functional stay-green phenotype.

Hierarchical cluster (hclust) analysis is a widely adopted clustering strategy where groups are assigned based on a dissimilarity measure supplied for all observations. Hierarchical clustering is most useful where the ideal number of clusters cannot be decided at the start of clustering analysis. In RNA-Seq analysis, it is more widely used for sample clustering, but there are significant implementations for gene clustering. Nose *et al.* (2020) clustered differentially expressed genes using the hclust function of masigPro, thereby highlighting the effects of day length and temperature on genes related to growth and starch metabolism in Japanese cedar (*Cryptomeria japonica* D. *Don*).

"mclust" is a powerful and popular R package, which allows modeling of data as a Gaussian finite mixture with different covariance structures and different numbers of mixture components for a variety of analyses (Scrucca *et al.*, 2016). It has been used in a broad range of contexts, including DNA sequence analysis (Verbist *et al*., 2015) and gene expression data (Nagoshi *et al*., 2004; Fraley & Raftery, 2006). Coolen *et al.* (2016) in monitoring the dynamics of co-expressed gene clusters in Arabidopsis, highlighted specific mclust-generated clusters and biological processes of which the timing of activation or repression was altered by prior stress.

Growth-Defense Balance in Plants

Comparative analysis of Arabidopsis transcriptional changes in response to treatment with *Pst* and its nonpathogenic strains at different time points has revealed that defense responses are initiated before pathogen multiplication (Mishina, 2007; Lewis *et al.*, 2015). The early-induced genes are related to defense responses and salicylic acid

(SA) biosynthesis (ref). In contrast, genes associated with photosynthesis-related processes are significantly suppressed, suggesting that plants may actively reduce the production of photosynthates to restrict the resources available for pathogen growth as an additional defense mechanism (Lewis *et al.*, 2015). Transcriptional changes of photosynthetic genes are of greater interest because the chloroplast is implicated in the defense response in several ways. Also, SA-dependent induction of pathogenesis-related (PR) genes and the HR is often dependent on light, but not functional chloroplasts (Goodman & Novacky, 1994; Dangl *et al.*, 1996; Griebel & Zeier, 2008). The growth-defense trade-off (Figure 2) has a significant impact on the agriculture industry since both processes are vital for increased plant fitness and thus has an influence on crop yields. A better understanding of this process could allow researchers to develop breeding strategies to optimize/maximize crop yield and meet the rising global food and biofuel demands. Therefore, I investigated the *cis*-regulatory control of transcriptional control in Arabidopsis following infection with virulent and avirulent strains of *Pst*.

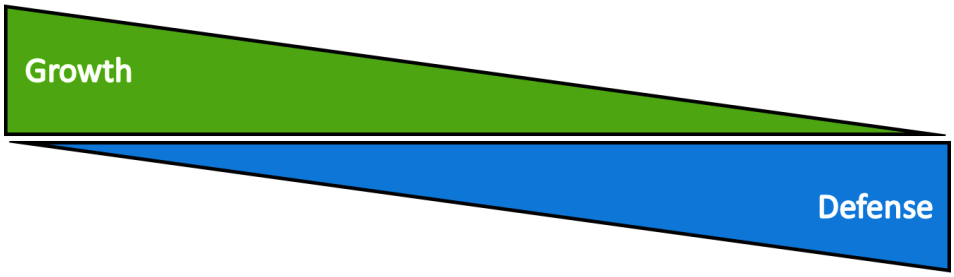


Figure 2. Illustration of Growth-Defense Trade-off in Plants. Growth and defense response to biotic and non-abiotic stressors are resource intensive processes. To maximize their chances of survival, plants prioritize defense response during pathogen infection, thus reducing the resources available for growth. Once defense response is complete, growth takes precedence and is referred to as the growth–defense tradeoff.

II. MATERIALS AND METHODS

Sample Preparation and Collection

To monitor transcriptome dynamics in Arabidopsis under biotic stress, Bordiya *et al.* (unpublished) performed RNA sequencing (RNA-Seq). This technique uses next-generation sequencing (NGS) to measure the presence and abundance of messenger RNA in a biological sample at a given moment. *Arabidopsis thaliana* L. ecotype Columbia (Col-0) plants were grown in soil at 22 °C, 60 % relative humidity, and at 16/8-hour light/dark cycle until sample collection. Three-and-half-week-old plants were syringe-infiltrated with 10mM $MgCl_2$ (Mock), *VirPst* ($5x10^5$ cfu/mL), or *AvrPst* ($5x10^5$ cfu/mL) using a needleless syringe, and harvested 1, 6, 24, and 48 hours post-infection (hpi) for preparing RNA. These treatments served to highlight differences in response to avirulent and virulent pathogen infection at early and latter time points. Additionally, uninfected plants (Naïve) and Mock treated plants served to reveal changes due to non-biotic disturbances during infection. Total RNA from the leaf tissue was extracted using PureLink RNA mini kit (Ambion). mRNAs were enriched with the polyA mRNA Magnetic Isolation Module (NEB). According to the manufacturer's protocol, enriched RNAs were reverse transcribed, end-repaired, and adapter-ligated using an RNA library preparation kit (NEB). Adapter ligation was performed with sample-specific barcodes. RNA-seq libraries were sequenced using an Illumina HiSeq 4000 at the UT Austin Genomic Sequencing and Analysis Facility (GSAF).

RNA-Seq Data Processing

Raw single-end sequence reads (100 bp) were de-multiplexed and sequencing quality was examined using FastQC v0.11.7 (Andrews, 2010). Raw reads were trimmed

for adapters using the Cutadapt software v.2.4 (Martin, 2011). Trimmed sequences of at least 40 bp for each sample were aligned to the Arabidopsis Araport11 (Cheng *et al*., 2017) reference genome and transcriptome annotation using Hisat2 v2.1.0 (Kim *et al*., 2019) with the --dta option and all other parameters set to default. SAM output files were compressed to BAM format and sorted using Samtools v1.9 (Li *et al*., 2009). All mapping statistics were obtained using Samtools. Aligned reads for each sample were assembled and merged based on the loci to which they mapped using Stingtie v2.1.3b (Pertea *et al.*, 2016) using default parameters with the option to natively estimate transcript abundance in FPKM. Alternatively, the prepDE.py (python) script, provided by the authors of Stringtie, was used to extract read count information. Principal components were computed using prcomp from the R package Stats v3.6.2 (Team & Worldwide, 2002). The first 3 principal components were visualized using the R package Scatterplot3D v0.3-41 (Ligges & Mächler, 2002).

Differential Expression Analysis

Identification of differentially expressed genes between select conditions was achieved using the R package edgeR v3.30.3 (Robinson *et al*., 2010) with a cut-off criterion of $0.5 > \log2(\text{fold change})$ or $\log2(\text{fold change}) > 2$; and Benjamini-Hochberg false discovery rate (FDR) < 0.05. edgeR employs the “Trimmed Mean of M-values” (TMM) normalization of libraries, which is recommended for RNA-Seq data where the majority (more than half) of the genes are believed to not be differentially expressed between any pair of the samples (Robinson & Oshlack, 2010). TMM normalized pseudocounts generated by edgeR were visualized as boxplots using the R Graphics v3.6.2 package (Murrelle, 2018). I tested the hypothesis that gene expression values of

VirPst and *AvrPst* infected plants are not equal to Mock expression values at any of the selected time points, after accounting for basal (Naïve) expression values. I also tested the hypothesis that expression values of *VirPst* and *AvrPst* infected plants are not equal. Counts of *VirPst* and *AvrPst*-induced DEGs per timepoint were visualized as bar plots using ggplot2 v.3.3.2 in R (Wickham, 2011).

Cluster Analysis of Genes Based on Transcriptional Dynamics

DEGs found in the differential expression analysis were pooled and clustered using three algorithms: the hclust and mclust options from the R package masigPro v1.60.0 (Nueda *et al*., 2014), and the SOM function from the kohonen v3.10.0 package (Wehrens & Kruisselbrink, 2018). For masigPro based clustering, a regression fit ($Q = 0.05$, and $\theta = 10$) was generated for all genes from which the best fit was selected using the T.fit function with default values. Significant genes ($R^2 = 0.5$) were extracted using the get.siggenes function. Significant genes were then clustered with the see.genes function cluster.method options: “Mclust” with k.mclust set to TRUE, and “hclust” with k set to 9. For kohonen, the expression matrix was scaled along genes, allowing a focus on the differences in the expression patterns instead of expression magnitude, and clustered into 9 hexagonal units using the SOM function. All gene expression profile visualizations were done using ggplot2. I visually examined individual genes in each SOM cluster, selecting for genes that showed high conformance to the mean expression profile for all genes in their SOM-assigned clusters.

GO Enrichment Analysis of DEGs in Up- and Down-regulated Clusters

Using the elim algorithm with fisher test statistic, topGO package in R (Alexa & Rahnenführer, 2009) was used to functionally annotate each cluster against the BioMart

database (www.plants.ensembl.org). The top 25 functional GO terms sorted by *p*-values for up- and down-regulated clusters were visualized using ggplot2.

Promoter Motif Analysis of Genes in Up- and Down-Regulated Clusters

An upstream sequence in 1kb from the transcriptional start site of all DEGs were collected. The collected sequences were sorted into their gene clusters and analyzed using MEME suite v5.1.1 tools (Bailey *et al*., 2009). Using MEME, the motifs appearing in most, but not all, sequences (ZOOPS) were collected. Motifs 5 to 25 bp in length, with an E-value less than 0.05, were selected. To remove false positives, all sequences collected were reshuffled and run through MEME using identical parameters.

Prediction of Binding Factors for Promoter Motifs in Up- and Down-regulated Defense Genes

Using the AME (Mc-Leay & Bailey, 2010; Franco-Zorrilla *et al*., 2014) implementation in the MEME suite, I tested for the enrichment of known motifs from the JASPAR Core Plants (2018) database (Khan *et al*., 2018), with one-tailed Fisher's exact test. E-value was computed as the *p*-value multiplied by the number of motifs in the input with all other parameters were set to default. Only motifs with enrichment E-values no greater 10 were be reported.

III. RESULTS

General Statistics and Features of RNA-Seq Data

The Bordiya *et al*. (unpublished) RNA-seq dataset totalled approximately 260 million single-end raw reads across all 39 libraries, with an average of about 6.7 million reads per library (Table 1). I confirmed adequate sequencing quality using FastQC; average QC_{30} was 38. Raw reads were then trimmed using Cutadapt to remove the TruSeq Indexed adaptor sequences, discarding reads in 40 bases or shorter (Martin, 2011). I obtained an average of about 6.6 million clean reads per library (Table 1). The trimmed reads were then mapped to the most updated Arabidopsis genome sequence, the Araport11 version, using Hisat2 with pre-defined default parameters (Kim *et al.*, 2019). The Hisat2 "--dta" option was used to tailor reported alignments for downstream transcript assembly. I obtained an average of about 6.2 million uniquely mapped reads per library (Table 1). For mapping statistics and further analysis of the alignment files, Samtools was employed.

Table 1. RNA-Seq mapping statistics for Naïve, Mock, VirPst, and AvrPst treated samples, 0, 1, 6, 24, and 48 hpi.

Sample	Reads Total	Trimmed Reads	% Trimmed Reads	Uniquely Mapped Reads	%Uniquely Mapped Reads
Naïve 0H 1	5,812,201	5,802,980	99.8	5,344,584	92.1
Naïve 0H 2	3,415,509	3,401,414	99.6	3,158,733	92.9
Naïve 0H 3	4,983,229	4,967,408	99.7	4,597,229	92.5
Mock 1H 1	5,875,430	5,871,964	99.9	5,531,159	94.2
Mock 1H 2	4,085,346	4,082,196	99.9	3,807,495	93.3
Mock 1H 3	3,776,990	3,761,627	99.6	3,516,162	93.5
Mock 6H 1	3,652,154	3,645,568	99.8	3,382,596	92.8
Mock 6H 2	5,388,944	5,380,253	99.8	5,007,911	93.1
Mock 6H 3	6,956,521	6,939,067	99.7	6,488,710	93.5

Mock 24H 1	4,144,931	4,135,494	99.8	3,868,920	93.6
Mock 24H 2	14,684,730	11,948,208	81.4	11,114,939	93.0
Mock 24H 3	1,752,481	1,749,406	99.8	1,628,793	93.1
Mock 48H 1	3,726,323	3,718,673	99.8	3,419,213	91.9
Mock 48H 2	7,505,290	7,479,182	99.7	6,883,290	92.0
Mock 48H 3	5,102,323	5,067,416	99.3	4,714,937	93.0
VirPst 1H 1	5,315,963	5,289,047	99.5	4,962,448	93.8
VirPst 1H 2	8,600,003	8,591,597	99.9	8,065,883	93.9
VirPst 1H 3	3,023,770	3,019,798	99.9	2,818,531	93.3
VirPst 6H 1	3,723,865	3,721,826	99.9	3,505,309	94.2
VirPst 6H 2	3,030,673	3,026,907	99.9	2,833,554	93.6
VirPst 6H 3	3,597,449	3,550,066	98.7	3,302,347	93.0
VirPst 24H 1	2,233,772	2,231,424	99.9	2,098,356	94.0
VirPst 24H 2	6,381,613	6,274,089	98.3	5,894,486	93.9
VirPst 24H 3	5,487,959	5,473,516	99.7	5,144,169	94.0
VirPst 48H 1	17,393,736	17,364,187	99.8	16,400,392	94.4
VirPst 48H 2	8,013,363	7,977,830	99.6	7,448,915	93.4
VirPst 48H 3	8,349,134	8,341,808	99.9	7,813,445	93.7
AvrPst 1H 1	6,580,425	6,568,656	99.8	6,175,938	94.0
AvrPst 1H 2	7,123,224	7,104,395	99.7	6,643,894	93.5
AvrPst 1H 3	7,173,509	7,158,596	99.8	6,725,074	93.9
AvrPst 6H 1	8,688,143	8,677,834	99.9	8,151,549	93.9
AvrPst 6H 2	21,546,807	21,511,849	99.8	20,257,452	94.2
AvrPst 6H 3	2,900,782	2,895,162	99.8	2,653,631	91.7
AvrPst 24H 1	1,518,056	1,516,482	99.9	1,382,251	91.1
AvrPst 24H 2	11,692,560	11,660,692	99.7	10,946,863	93.9
AvrPst 24H 3	9,217,883	9,211,194	99.9	8,720,200	94.7
AvrPst 48H 1	1,675,810	1,673,572	99.9	1,552,306	92.8
AvrPst 48H 2	14,506,721	14,494,538	99.9	13,631,731	94.0
AvrPst 48H 3	11,626,677	11,621,795	100.0	10,868,194	93.5
Total	**260,264,299**	**256,907,716**		**240,461,589**	

Sample ID contains treatment, hpi, and replicate information. Uniquely mapped reads are mapped to only one region of the Arabidopsis genome.

I assembled and merged reads aligned to distinct gene loci within the Arabidopsis genome using Stringtie with default parameters (Pertea *et al*., 2015). Estimates of transcript abundances (in FPKM) and counts tables were natively generated by Stringtie.

For the comparison of the expression value distribution, normalized counts were visualized as boxplots using R Graphics (Figure 3). To examine the level of similarity/dissimilarity in the gene expression profiles of the different treatments and time points, a principal component analysis (PCA) was performed using the Stats package and visualized as 3D plots in R. (Figure 4).

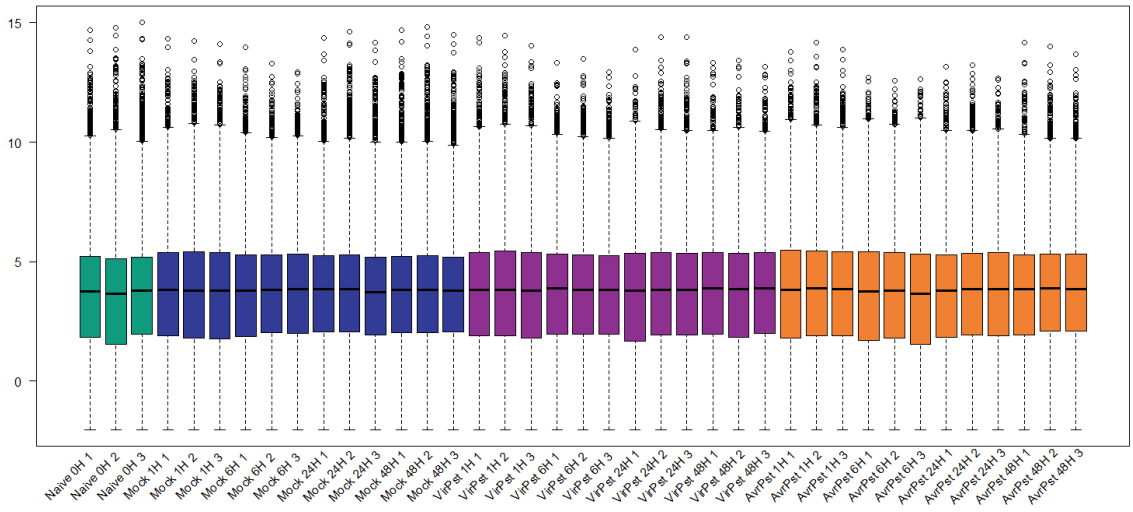


Figure 3. Distribution of Normalize Reads in the RNA-seq Datasets Using Box Plot Representation. Box plots representing log 2 TMM normalized gene expression distributions for all CDS in Naïve (green), Mock (blue), *VirPst* (purple), and *AvrPst* (orange) treated samples. All samples showed a comparable range of gene expression.

Genes are Responsive to *VirPst* and *AvrPst* Treatments

I applied a count-based negative binomial model implemented in the R package "edgeR", to identify genes that are differentially expressed during the plant biotic stress response (Robinson *et al.*, 2010). I compared gene expression between *VirPst* and Mock, *AvrPst* and Mock, and *VirPst* and *AvrPst*, identifying 2,613, 3,062, and 2,766 differentially expressed genes (DEGs) (adjusted-$p < 0.05$), respectively. In total, 4,306 DEGs were found, out of which 2,310 were up-regulated while 2,038 were down-

regulated at any of the sampled time points. Amongst these DEGs 1, 3,166, 2,151, and 1,776 were differentially expressed at 1, 6, 24 and 48 hpi, respectively (Figure 5). The majority of the genes up- or down-regulated in response to *VirPst* treatment were also up-regulated or down-regulated by *AvrPst* treatment, while others were unique to each treatment (Figure 6).

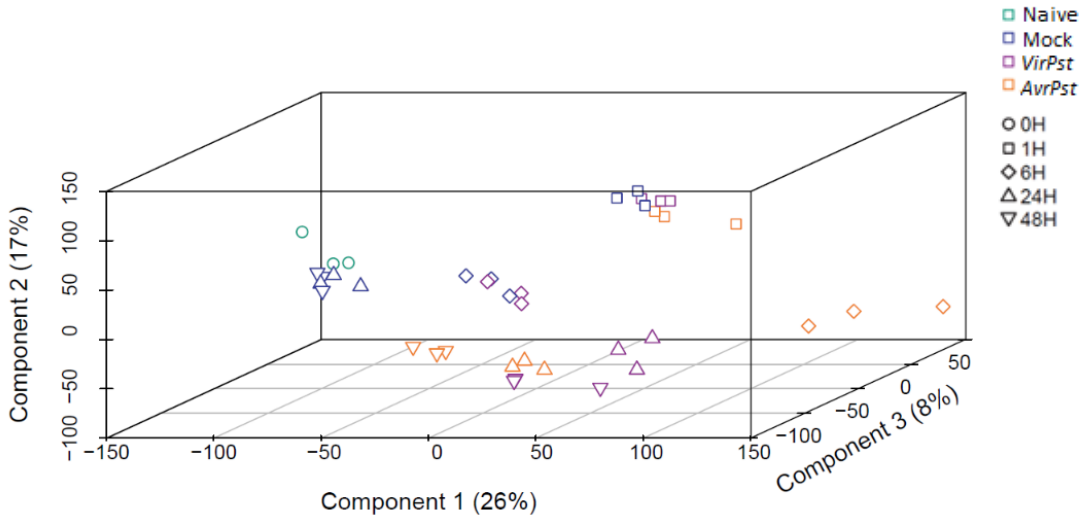


Figure 4. Principal Component Analysis Showing Correlation Among Different Treatments. Three-dimensional PCA scatter plot of FPKM normalized gene expression in Naïve (green), Mock (blue), *VirPst* (purple), and *AvrPst* (orange) treated samples at 0, 1, 6, 24, and 48 hpi. The percentages on each axis represent the percentages of variation explained by the first, second, and third principal components, respectively.

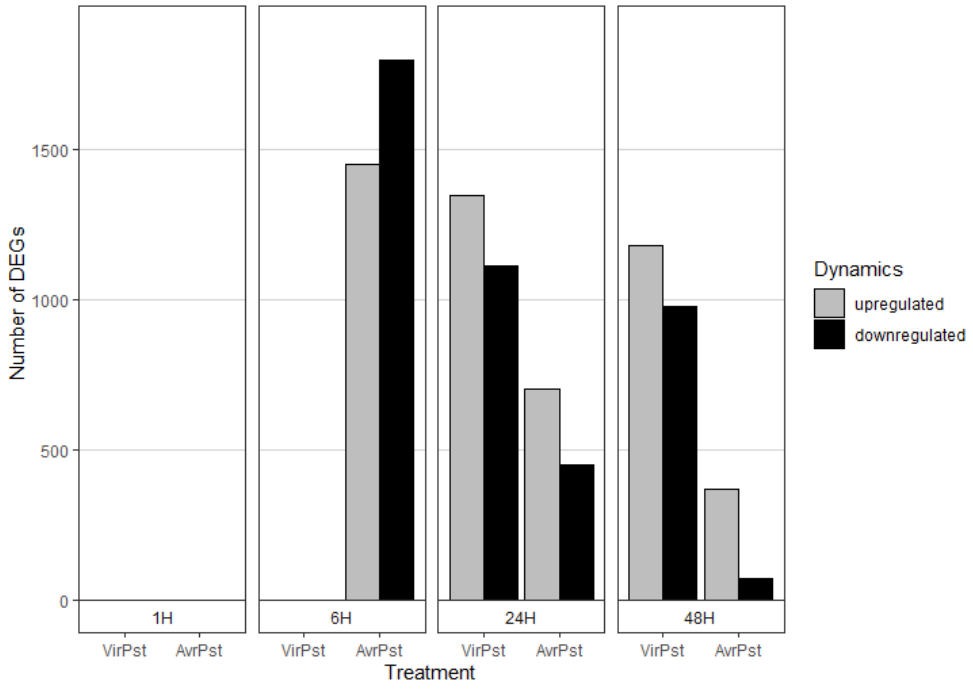


Figure 5. Bar Plot Showing Counts of DEGs at 1, 6, 24, and 48 Hours Post-infection with Mock, *VirPst,* and *AvrPst*. Up-regulated genes (grey bar) are significantly induced (fold change >= 2; FDR < 0.05) while down-regulated genes (black bars) are significantly repressed (fold change <=0.5; FDR < 0.05) in pathogen treated samples as compared to Naïve and Mock treated samples.

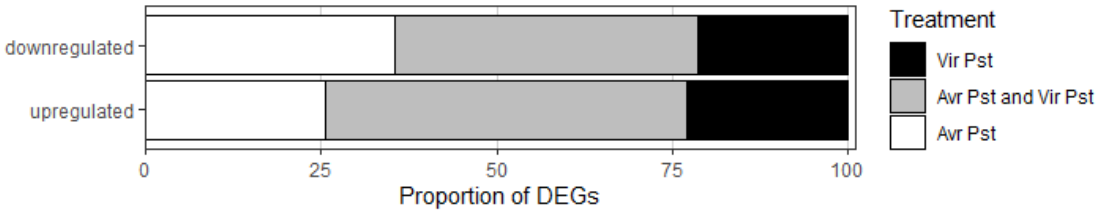


Figure 6. Bar Plot Showing Overlaps of DEGs between Pathogen Treatments. *VirPst* induced/repressed DEGs (black) overlap with *AvrPst* induced/repressed DEGs (white) and the overlaps are represented by the grey sections.

Defense Genes Displaying Dynamic Transcriptional Changes

I decided to group DEGs based on the similarity of their expression patterns in response to biotic stress using unsupervised clustering. To evaluate each clustering strategy for our data set, I tested three clustering strategies as implemented in masigPro and kohonen R packages. Using model-based clustering (mclust), hierarchical clustering (hclust), and self-organized maps (SOM), I obtained 8, 9, and 9 distinct gene clusters, respectively. To examine the reliability of these clustering algorithms, I further resolved each cluster generated by the algorithms into sub-clusters; i.e., another clustering round of each cluster was performed to see if a different pattern emerges. I define accuracy here as the tendency of clusters from each algorithm to maintain their representative expression patterns upon sub-clustering. I obtained 77 %, 58 %, and 91 % accuracy from mclust, hclust, and SOM, respectively (Figure 8). Our results suggest that SOM was the most reliable algorithm for clustering our RNA-seq data.

Five out of the nine SOM clusters comprised genes up-regulated (U1, U2, U3, U4, and U5) to different extents by the treatments. The four remaining clusters were down-regulated (D1, D2, D3, and D4; Figure 9). The nine clusters of DEGs showed the following characteristics:

U1. minimum induction by Mock at 6 hpi; rapid but brief induction by *AvrPst* at 6 hpi; a gradual and long-lasting induction by *VirPst* till 48 hpi;

U2. brief suppression at 1 hpi by all treatment; marginal induction around two-fold by Mock treatment at 6 hpi followed by gradual decrease; comparable induction by *AvrPst* and *VirPst* at 6 hpi, followed by suppression by *AvrPst* while further increase by *VirPst*;

U3. brief and notable induction by all treatments at 1 hpi; rapid but brief induction by *AvrPst* at 6 hpi; gradual induction by *AvrPst* peaking at 24 hpi;

U4. marginal induction by Mock at 1 hpi; gradual and sustained induction by *AvrPst* peaking at 24 hpi; gradual and sustained induction by *VirPst* treatment till 48 hpi;

U5. significant induction by all treatments at 1 hpi; basal level afterward by Mock; gradual decrease till 48 hpi by *AvrPst*; rapid reduction to basal level at 6 hpi, followed by another peak at 24 hpi by *VirPst*;

D1. repressed by all treatments at 1 hpi; further suppression by *AvrPst* till 6 hpi; while returning to the pre-treatment level, marginal induction at 6 hpi by *VirPst* and Mock; additional suppression by *VirPst* at 24 hpi;

D2. repressed by all treatments at 1 hpi; further suppression by *AvrPst* till 6 hpi; return to the pre-treatment level by *VirPst* and Mock while additional suppression by *VirPst* at 24 hpi;

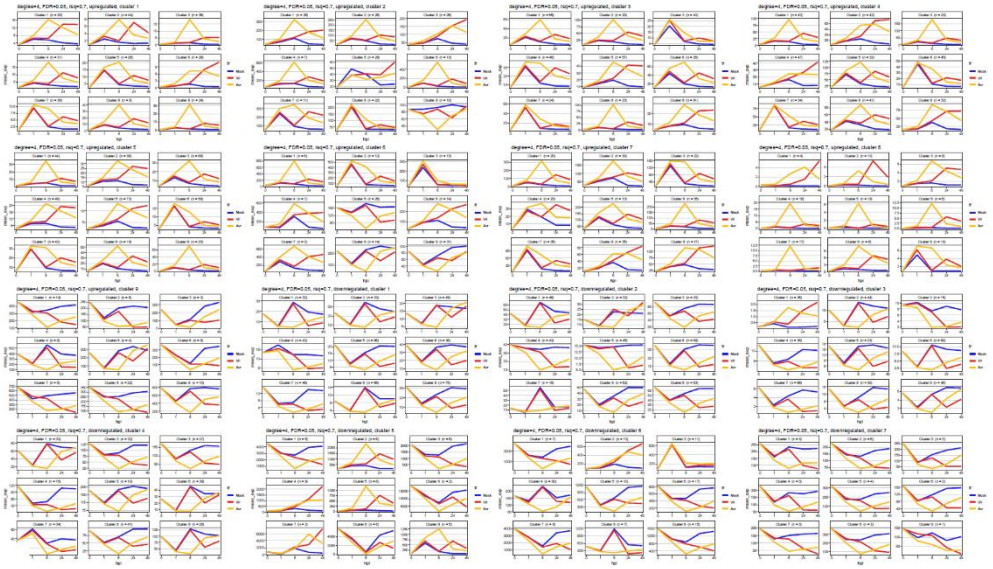
D3. minimally induced by all treatments at 1 hpi; return to the pre-treatment level by 6 hpi by Mock; rapid suppression at 6 hpi and gradual recovery by *AvrPst*; gradual sustained suppression at 24 and 48 hpi; and

D4. highly expressed genes before treatment; most rapidly repressed by *AvrPst* peaking at 6 hpi; marginal decrease at 6 hpi by Mock; gradual and sustained decrease by *VirPst* till 48 hpi.

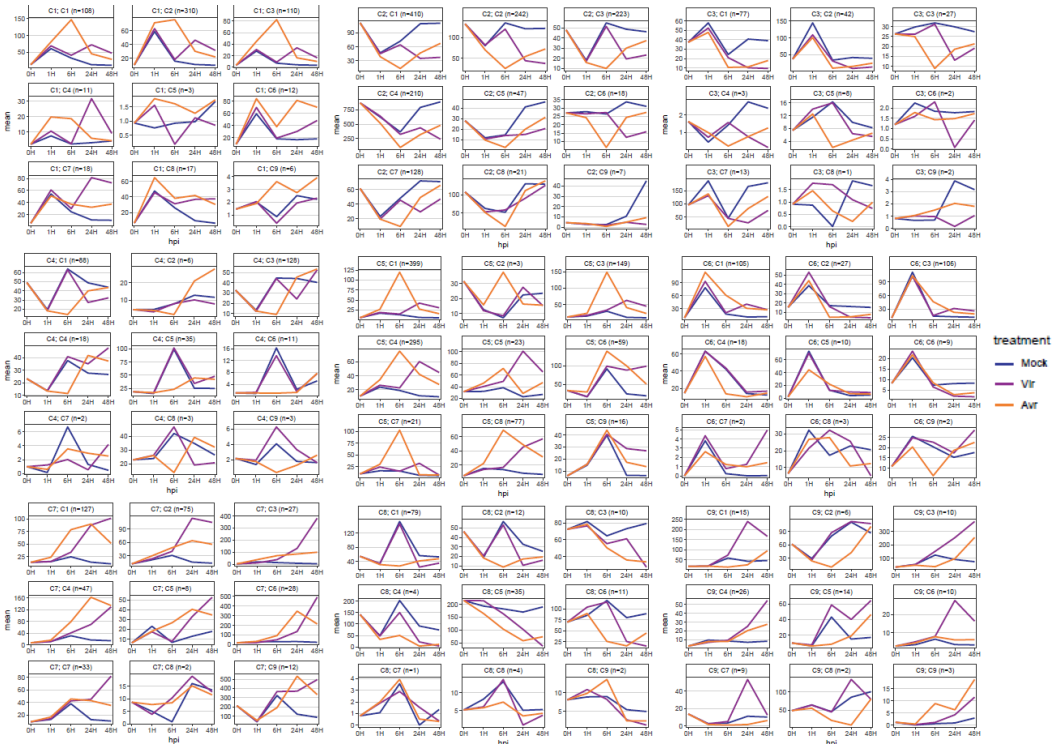
To further improve confidence in the clusters, I manually examined expression profiles for the clustered genes. I filtered for genes whose expression profiles mimicked the expression profile of their SOM-assigned cluster (Figure 10). The high-conforming genes within each cluster were superimposed and visualized.

To functionally categorize our gene clusters, I performed GO analysis using the topGo package in R. The most significantly enriched GO terms in the up-regulated clusters are stress response-related (Figure 11). For the down-regulated gene clusters, the most significantly enriched GO term was photosynthesis.

A.



B.



C.

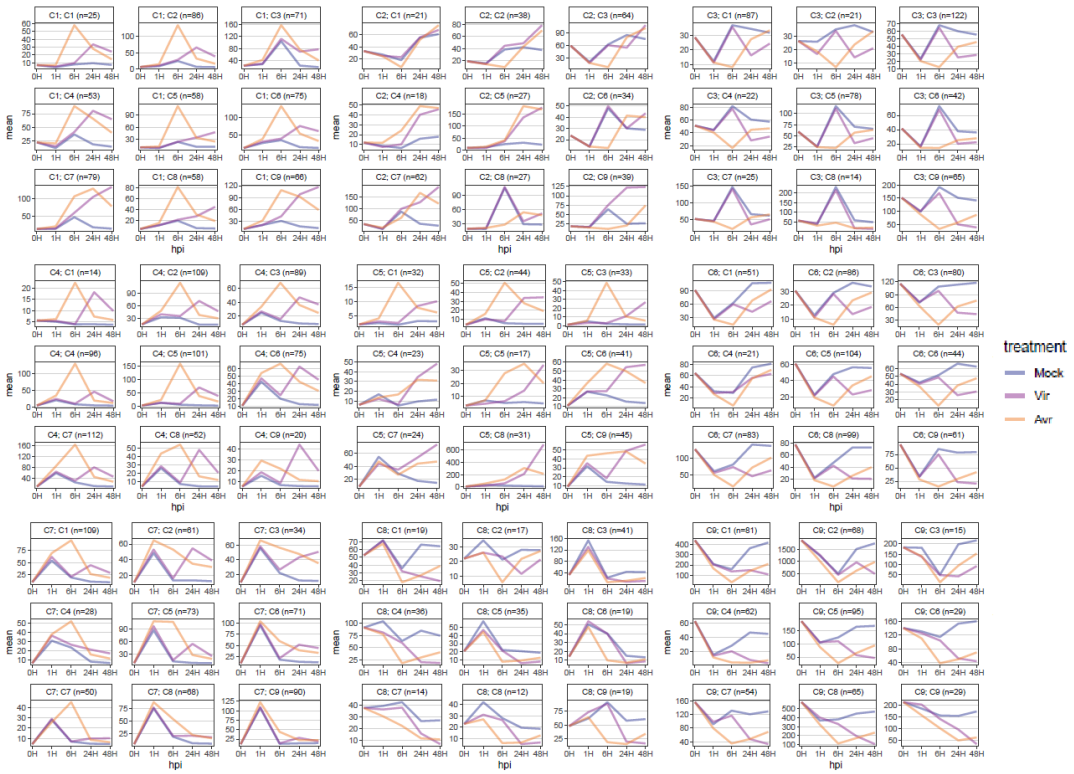


Figure 7. Test Accuracy of Clustering Algorithms by Sub-clustering. (A) hclust, (B) mclust, and (C) SOM generated clusters were further resolved, resulting in 9 clusters for each cluster. Y-axes represent mean feature kilobase per million reads mapped (FPKM) within sub-clusters, whereas X-axes represent hpi. Colored lines depict the mean expression profiles for each treatment within the clusters.

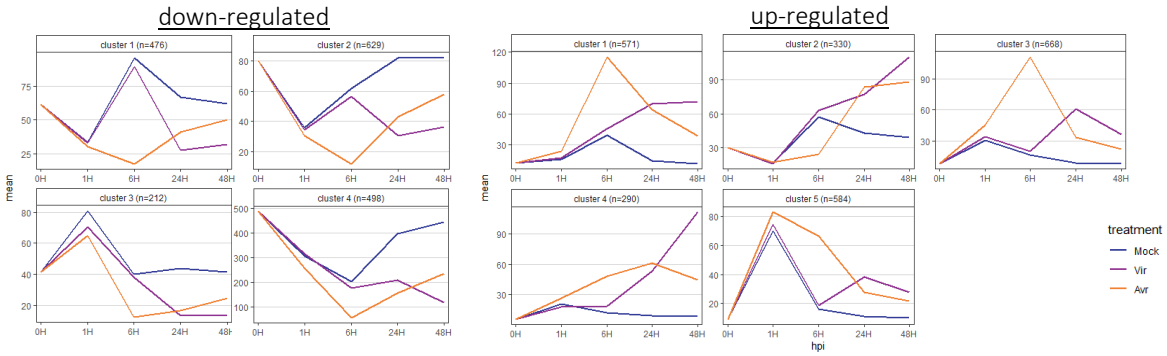


Figure 8. Contrasting Gene Expression Profiles of SOM Clusters at 0, 1, 6, 24, and 48 hpi. Y-axes represent mean feature kilobase per million reads mapped (FPKM) within each cluster, whereas horizontal axes represent hpi. Colored lines depict the expression profiles for each treatment within the clusters.

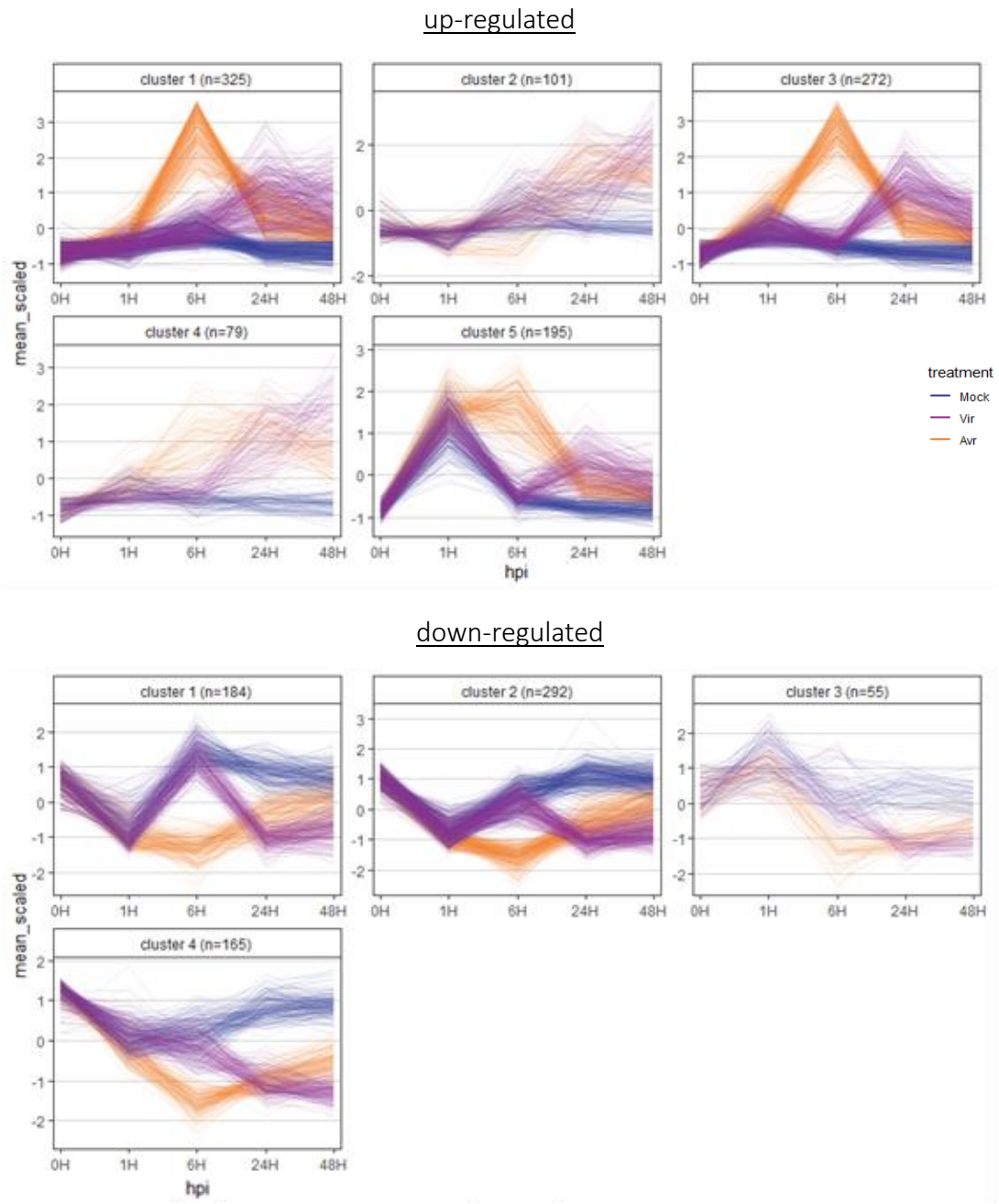


Figure 9. Clustering Improved by Manual Inspection. Member genes of each SOM cluster were manually checked to ensure that they were assigned to the right clusters. Poorly conforming genes were discarded, and the surviving genes plotted individually. Y-axes represents feature kilobase per million reads mapped (FPKM), whereas horizontal axes represent hpi. Colored lines depict the expression profiles for each treatment within the clusters.

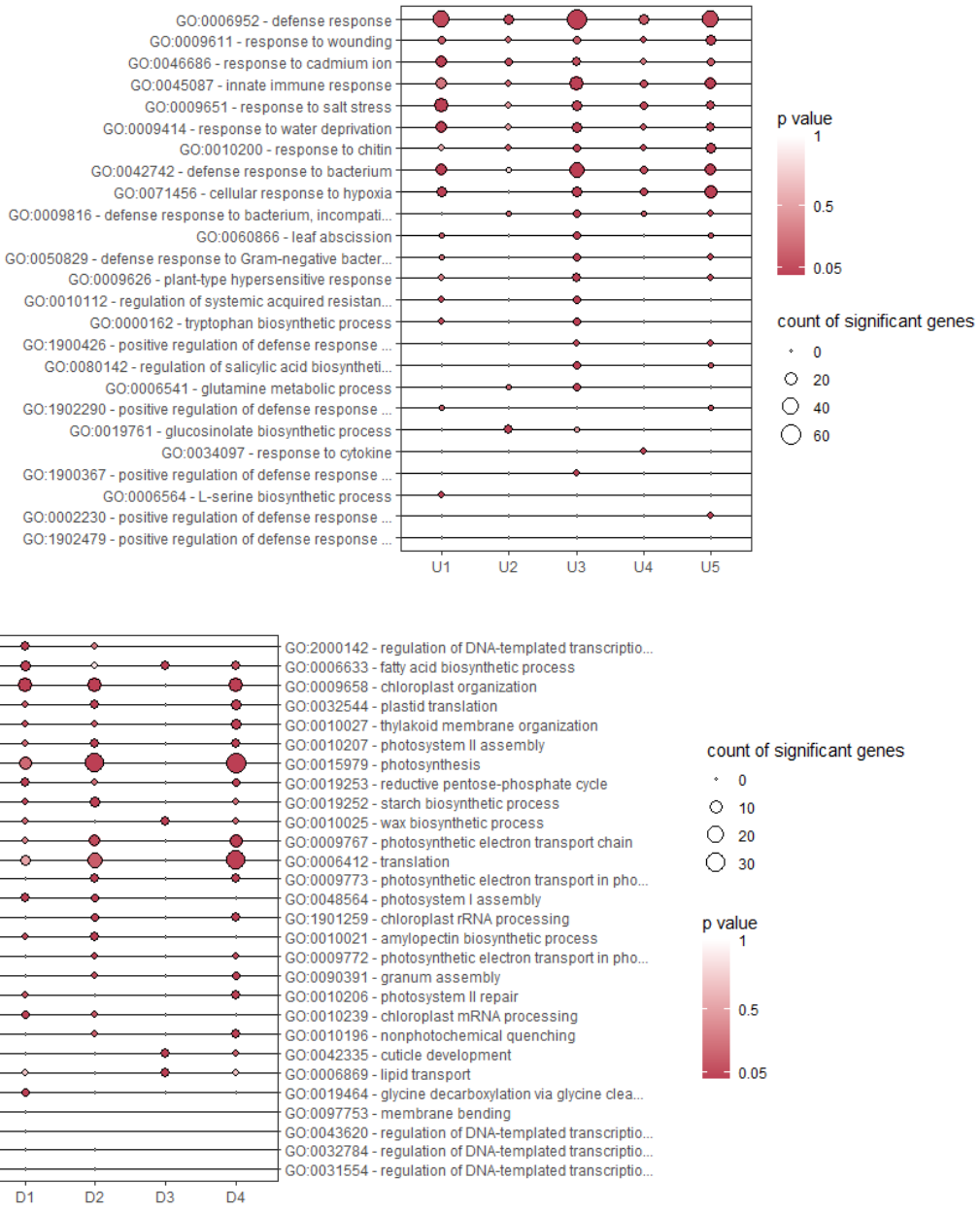


Figure 10. Dot Plot of Enriched GO Terms in each Cluster. GO enrichment analysis of DEGs was retrieved using topGO. The top 25 most significantly ($p < 0.05$) enriched GO terms in biological process are plotted in descending order of the significant gene number. The size of the dots represents the number of DEGs enriched in each GO term in each cluster and the color of the dots represent the adjusted *p*-values.

Binding Motifs of Transcription Factor (TF) Binding Sites

I sought to provide insight into the TF-mediated control of gene expression during biotic stress challenge to Arabidopsis by performing a detailed analysis of *cis*-elements within promoter regions of each expression cluster. For each cluster, I analyzed promoter sequences, 1 kb upstream of TSS, of associated genes for common motifs, using the MEME tool. I found 4, 3, 3, 3, and 5 significant motifs in U1, U2, U3, U4, and U5 clusters, respectively (Figure 12). For the down-regulated clusters, I found 4, 4, 3, and 5 significantly enriched motifs in D1, D2, D3, and D4 clusters, respectively.

I further enriched for known transcription factor binding sites in the JASPAR Core Plants (2018) database, using the AME tool from MEME suite. TF binding sites with an E-value not less than 10 were selected as known TF binding sites enriched in each cluster (Table 2). These TF binding sites were sorted by their likelihood to be false positives, which is given as the percentage of motifs labeled as positive (Pos) and classified as a true positive (TP). These results were visualized as dot plots, with the color of the dots representing the motif binding affinity calculated as *p*-values in $\log_{10}$.

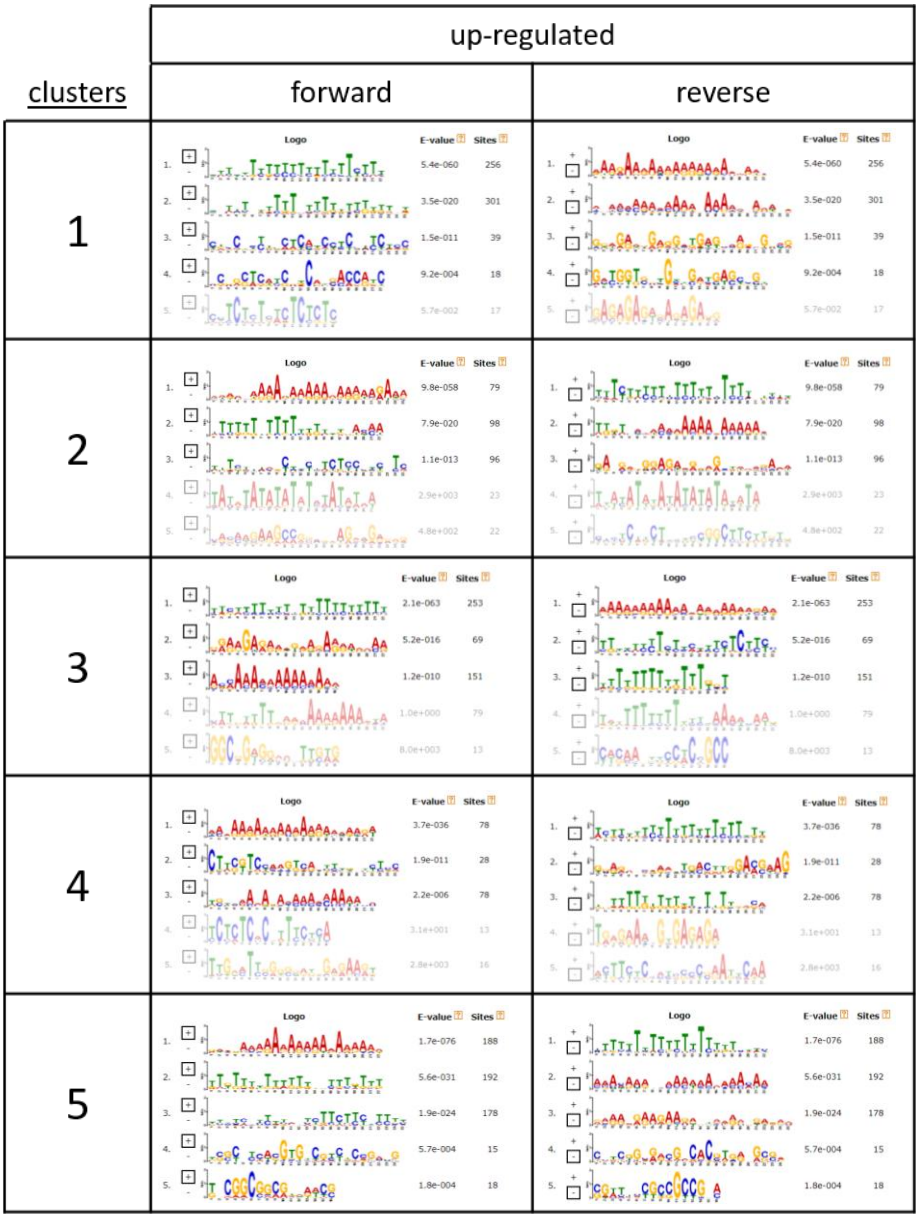


Figure 11. Comparison of Motifs between Up-regulated Clusters. Motif were identified in the promoters of genes within each up-regulated cluster, allowing for some promoters not to have every motif found. Top 5 motifs that are 5 to 25 bp in length, with an E-value less than 0.05, and are not found in shuffled (control) sequences are displayed.

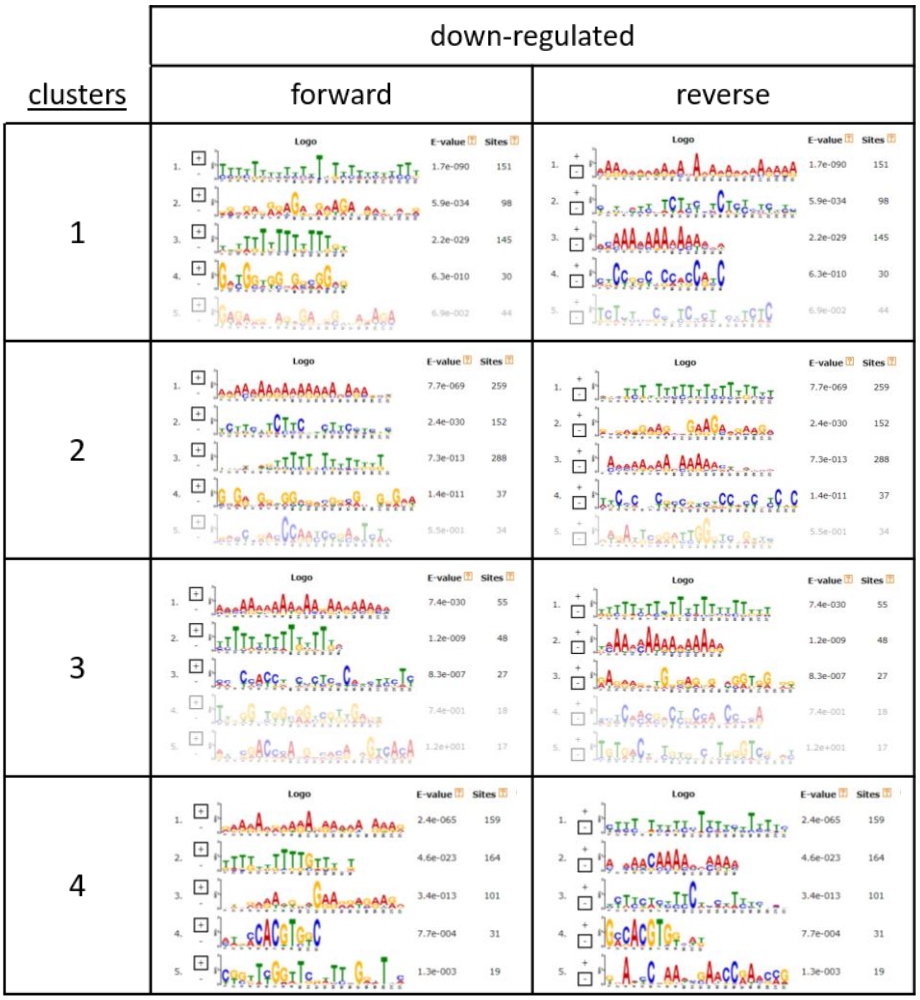


Figure 12. Comparison of Motifs between Down-regulated Clusters. Motifs were identified in the promoters of genes within each down-regulated cluster, allowing for some promoters not to have every motif found. Top 5 motifs that are 5 to 25 bp in length, with an E-value less than 0.05, and are not found in shuffled (control) sequences are displayed.

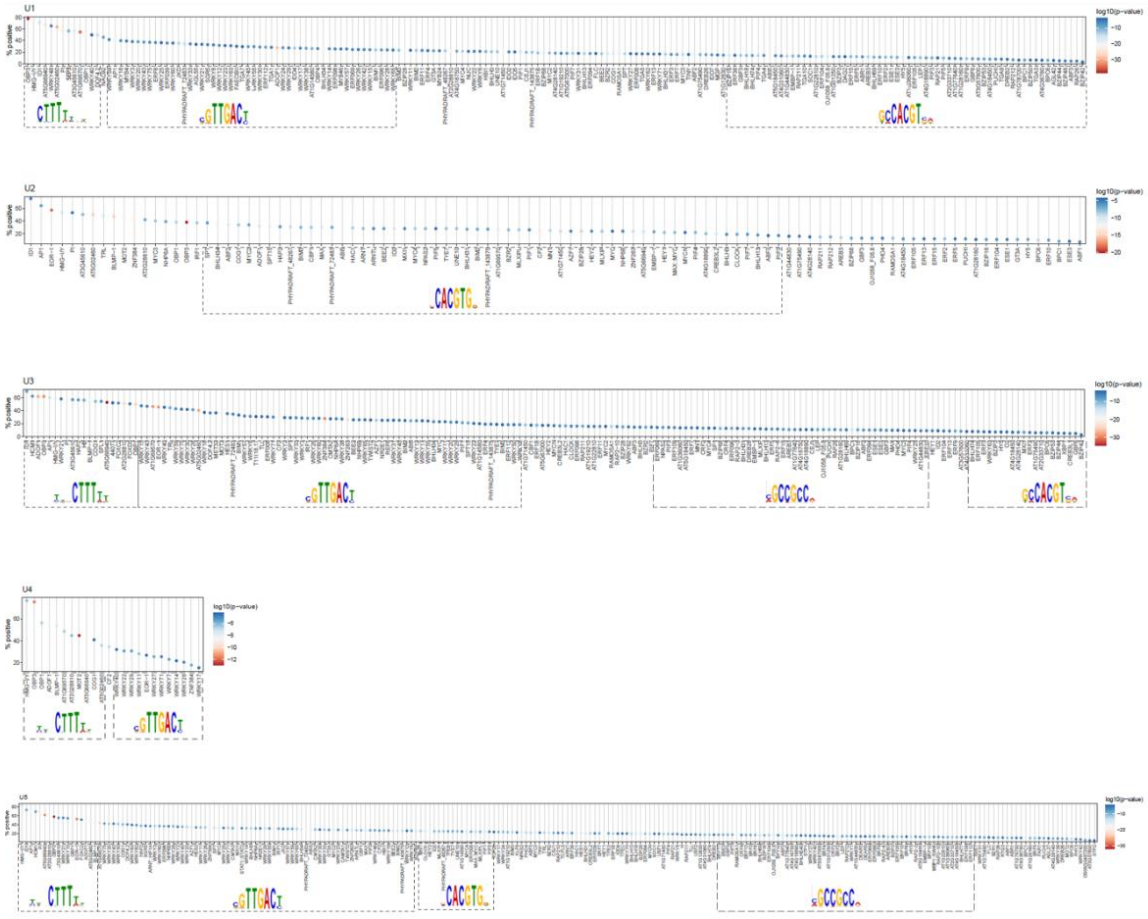


Figure 13. Comparison of known Transcription Factor Binding Sites between Up-regulated Clusters. Known TF sites were enriched in the promoters of genes within each up-regulated cluster. Enriched binding sites with E-value greater than 10 are displayed as dot plots. Binding sites are sorted by their % TP values and sequence logos shown for binding site motifs that clustered together.

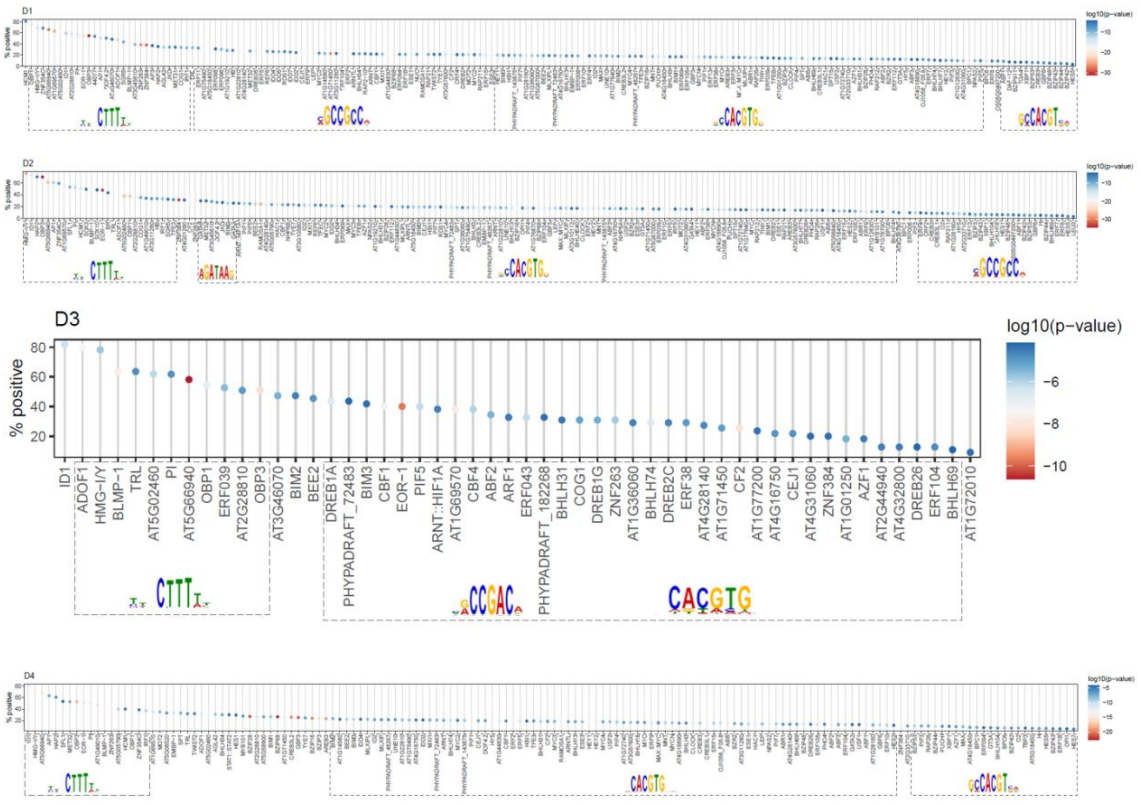


Figure 14. Comparison of known Transcription Factor Binding Sites between Down-regulated Clusters. Known TF sites were enriched in the promoters of genes within each down-regulated cluster. Enriched binding sites with E-value greater than 10 are displayed as dot plots. Binding sites are sorted by their % TP values and sequence logos shown for binding site motifs that clustered together.

IV. DISCUSSION

I analyzed the transcriptome dynamics of wildtype *Arabidopsis thaliana* responding to *AvrPst* and *VirPst* infection at 0, 1, 6, 24 and 48 hpi. I identified 4,306 DEGs in total, with 2,204 and 2,535 being up- and down-regulated, respectively, by pathogen treatments at any of the sampled time points. These DEGs were clustered by their gene expression patterns into 9 gene groups. These 9 groups were divided to either up-regulated and down-regulated. Up-regulated clusters had a total of 2,443 genes which display induction of gene expression in pathogen treated samples as compared to Mock treated samples. Down-regulated clusters comprised 1,815 genes characterized by a repression of gene expression in pathogen treated samples as compared to Mock treated samples. Our analyses of promoter sequences of genes within each cluster revealed the presence of known functional motifs/TF binding sites that correlate with the observed expression data. I observed an enrichment of the core W-box sequence (WRKY binding sites), which are known regulators of defense signaling (Eulgem *et al*., 2000; Euglem & Somssich, 2007), and the G-box sequence (bHLH/bZIP binding sites), positive and negative regulators of photosynthesis (Kindgren *et al*., 2012; Chattopadhyay *et al*., 1998; Leivar & Quail, 2011; Leivar & Monte, 2014; Castillon *et al*., 2007), in up-regulated and down-regulated gene clusters, respectively. I also found the P-box core (CTTTT), which is bound by Dof proteins (Yanagisawa 2002), in all 9 gene clusters. Together, I believe that the enrichment of these TF binding sites in the respective clusters, accounts at least in part, for the induction repression patters observed in this study.

Transcriptome Dynamics for Defense Genes

Transcriptome dynamic analysis of Arabidopsis infection with *VirPst* and *AvrPst* revealed diverse patterns (Figures 9 and 10). These dynamic patterns were largely grouped into either up-regulated or down-regulated relative to the pre-treatment level. Most up-regulated (U) clusters except for U2 contained genes with very little basal expression, suggesting that these genes are tightly suppressed under normal growth conditions.

U1 and U3 clusters were rapidly induced by *AvrPst*, and their upregulation peaked at 6 hpi (Figures 9 and 10). Among the up-regulated defense genes, a majority of these defense genes peaked at 6 hpi with *AvrPst* (Figure 9). It has been shown that coronatine, a phytotoxin derived from *Pst*, is delivered by 6 hpi or earlier, suggesting that pathogen-derived virulence factors would be detectable for R proteins before 6 hpi (de Torres Zabala *et al*., 2009). Consistent with this timeline, another kinetic study investigating the transcriptome dynamics in response to *AvrPst* also found that 6 hpi is the major time point at which most of the defense genes appear to peak (Mine *et al*., 2018). Thus, these observations together with mine suggest that reprogramming transcriptome specific to ETI likely occurs before 6 hpi.

The induction kinetics in Arabidopsis infected with *VirPst* was significantly slowed relative to the ETI counterpart; the induction peak was delayed to 24 hpi or later. The delayed induction by *VirPst* likely is contributed from its battery of virulence factors, including effectors that are known to smother PTI-associated transcriptional reprogramming (Lewis *et al*., 2015). However, to accurately assess the impact of the virulence factors, I would have needed a transcriptome dataset for PTI. Note that the

transcriptome data obtained from the *VirPst* treatment in this study provides information on basal resistance. In essence, this basal resistance is the remaining defense in which PTI was negated by the pathogen virulence factors. Thus, the difference between *VirPst* and *AvrPst* in U1, U3, and U4 clusters may instead be attributed to an R protein initiating ETI.

Plants' ability to recognize a pathogen effector (a would-be virulence factor) by an R protein(s) that induce ETI frequently determines whether the outcome of infection results in severe disease (Jones and Dangl, 2006). Interestingly, while ETI is known to be much stronger than PTI, the difference is mostly quantitative rather than qualitative (Tao *et al*., 2003, Tsuda *et al*., 2009), involving largely the same set of defense genes. This notion of the quantitative difference was confirmed in my analysis since 88 % of the induced genes by *AvrPst* in U1, U3, and U4 clusters were also induced but later by *VirPst* (Figure 6)

The down-regulated genes also showed a pattern comparable to the up-regulated clusters; i.e., more rapid and robust suppression in response to *AvrPst* relative to *VirPst*. In contrast to the inducible defense genes, the down-regulated genes are mostly constitutive genes, which were briefly repressed to near-zero expression levels at 6 hpi by *AvrPst* infection and much later at 48 hpi, albeit to a lesser degree by *VirPst* infection. In particular, D4 cluster, with very high basal expression level that are 5 to 10 times higher than the rest of the down-regulated clusters (Figure 9), has genes displaying rapid suppression and subsequent recovery, suggesting that this speedy transcriptome kinetics is important in ETI. As analyzed in Figure 11, this group of down-regulated defense genes is enriched for GO terms such as "photosynthesis", "chloroplast organization", and

"electron transport chain", all of which are all photosynthesis-related. Furthermore, my *cis*-element and binding factor analysis raised the possibility that this suppression could be mediated by the G-box and its binding factor PIF (Paik *et al*., 2017). This ETI-mediated suppression of photosynthetic genes observed in my study potentially provide a molecular mechanism underlying defense-growth trade-off (Figure 2), which has been reported for over a decade (Huot *et al*., 2014; Zou *et al*., 2005; Smakowska *et al*., 2016; Gerth *et al.*, 2017;).

Upon closer examination of the U4 gene cluster, I found that commonly used defense marker genes, *PR1* (AT2G14610), *PR2* (AT3G57260), and *PR5* (AT1G75040) are present. These same genes were found to be up-regulated throughout defense induction, that is, 6 to 24 hpi with *AvrPst* and 24 to 48 hpi with *VirPst* treatments. U4, however, is the least populous up-regulated cluster displaying rather slower induction-kinetics relative to U1, U3, and U5. My observation supports the need for marker genes representing the groups with early induction kinetics, which comprise most defense genes.

Cis-Elements and their Binding Factors Associated with Biotic Stress

I found a few DNA-binding motifs to be significantly enriched in promoter sequences of each gene cluster (Figure 12 & 13). Using AME, I found well-characterized cis-elements/ TF binding sites to be enriched in different clusters. I found the core W-box sequence, (T)TGAC(C/T), to be enriched in the promoters of U1, U3, U4, and U5 genes but not in down-regulated clusters. The binding of WRKY proteins to W-box is a well-characterized feature of biotic and abiotic stress response. Consistent with this finding, W-boxes have been found in the promoter sequences of various genes that responded to

wounding or pathogens, including the genes encoding *PR* proteins (Ülker and Somssich, 2004). Hence, I believe that the presence of these WRKY binding sites, is at least in part, responsible for the induction of up-regulated genes by *AvrPst* and *VirPst* treatments. Overrepresentation of W-box motifs may also explain the delayed induction in *VirPst* treatment, as expression of some WRKYs are specifically suppressed by virulent strains of *P. syringae* in an effector-dependent manner (Higashi *et al*., 2008).

Inversely, I found the core G-box sequence, (CACGTG), to be enriched in the promoters of all down-regulated gene clusters but not in early (U1 and U2) and late (U4) up-regulated clusters. The binding of basic leucine zipper (bZIP) and basic helix-loop-helix (bHLH) proteins to G-box is well-characterized (Katagiri & Chua, 1992; Kindgren *et al*., 2012; Castillon *et al*., 2007). A subgroup of bHLH, phytochrome-interacting factors (PIFs), have been shown to contribute to the growth-defense response in plants (Paik *et al*., 2017). In particular, PIF3 has been shown to suppress chloroplast development (Stephenson *et al*., 2009, Liu *et al*., 2017) and displays transcription repression as well as activation activities (Zhang *et al*., 2013). Therefore, it is feasible that increased PIF3 or its homolog in the nucleus in response to pathogen infection could suppress photosynthesis and thereby limit growth with the observed suppression kinetics. This possibility was confirmed by our recent outcomes: i) overexpression of *PIF3* enhanced bacterial resistance in Arabidopsis and ii) its nuclear level was increased in response to *AvrPst* (Nam & Kang, unpublished).

Coronatine (COR), a bacterial JA-Ile mimic (Feys *et al*., 1994), and several bacterial effectors have been shown to degrade JAZ repressor proteins (Gimenez-Ibanez *et al*., 2014; Yang *et al*., 2017; Jiang *et al*., 2013; Zhou *et al*., 2015), thus surpressing host

defense responses via JA signaling. Also, more studies have shown that in addition to the increase in endogenous SA levels during ETI, the JA concentration also spikes at 3 to 9 hours post-infection (Kenton *et al.*, 2007; Spoel *et al.*, 2007). In the presence of SA, NPR3 and NPR4 instigate the proteasome-mediated breakdown of JAZ proteins (Liu *et al.*, 2016). This scarcity of JAZ proteins frees up DELLAs (Hou *et al.*, 2010), which in turn dimerizes with PIFs and with other bHLH transcription factors (Gallego-Bartolomé *et al*., 2010; Li *et al*., 2017). This dimerization with DELLA repressors renders PIFs less active. In a recent study, DELLA proteins also were shown to promote the degradation of all four PIFs through the 26S proteasome pathway under both dark and light conditions (Li *et al*., 2016). This degradation and/or loss of ability of PIFs to bind to target genes likely play a regulatory role in modulating photosynthesis/growth although this mechaniam in decreasing PIF activities counteracts increase in nuclear PIF3 proteins as described above. Thus, the analysis of these regulatory factors over multiple time points will provide clear underlying regulatory mechanism for suppressing photosynthesis and growth.

Also, I found the "CTTT" motif to be enriched in the promoter sequences of all our clusters. (A/T)AAAG or its complementary sequence is usually found in the binding sequence of DNA-binding with one finger (Dof) proteins (Kang & Singh, 2001; Yanagisawa 2002). Dof proteins have been implicated in a host of functions ranging from plant defense gene expression, to hormone response, growth, amongst others. In some cases, closely related Dof proteins have been shown to play opposite roles in regulating gene expression. It is therefore not surprising that Dof binding sites were found in both up-regulated and down-regulated genes.

Clustering Algorithms

While many strategies have been proposed to group/cluster expression data from time-course data, there is no consensus to date on which method is most suitable for a given dataset. Spies *et al* (2019) attempted to compare several available algorithms and found splineT and maSigPro performed better than pairwise approaches with less false-positive candidates. I evaluated the reliability of three clustering strategies to generate clusters; two from maSigPro (Nueda *et al*., 2014) and one from kohonen (Wehrens & Buydens, 2007). The self-organized mapping-based method provided the best performance, with an accuracy of 91 % (Figure 7). In contrast, the hierarchical method yielded an accuracy of 58 %, the lowest of the three tested algorithms. Gaussian mixture modelling for model-based method fared slightly better with an accuracy of 77 %. As time-course RNA-seq becomes more common, the reliability of the clustering algorithms needs to be carefully scrutinized/evaluated, given the lower reproducibility of some clustering algorithms found in my study.

Reprogramming transcriptome is a crucial event in antibacterial defense responses (Mine *et al*., 2018). My time-course transcriptome analysis provided significant insight into how two different types of resistance, ETI and basal resistance, operate at the level of RNA transcription. Intriguingly, in addition to traditional defense genes with the induction kinetics, I uncovered a group of genes that are rapidly suppressed under biotic stress, suggesting that this novel type of defense genes also contributes to plant resistance. It has been observed that defense responses come with a fitness cost for over a decade (Burdon & Thrall, 2003). I propose that optimizing these defense genes with the

suppression kinetics would likely facilitate the development of resistance traits without yield loss in plants.

REFERENCES

Afzal, J., Williams, M., Leng, M. J., & Aldridge, R. J. (2011). Dynamic response of the shallow marine benthic ecosystem to regional and pan-Tethyan environmental change at the Paleocene–Eocene boundary. *Palaeogeography, Palaeoclimatology, Palaeoecology, 309*(3-4), 141-160.

Alexa, A., & Rahnenführer, J. (2009). Gene set enrichment analysis with topGO. URL https://rdrr.io/bioc/topGO/f/inst/doc/topGO.pdf.

Andrews, S. (2010). Babraham bioinformatics-FastQC a quality control tool for high throughput sequence data. URL https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... & Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids research, 37*(suppl_2), W202-W208.

Berkhin, P. (2006). A survey of clustering data mining techniques. In *Grouping multidimensional data* (pp. 25-71). Berlin, Heidelberg: Springer.

Blume, B., Nürnberger, T., Nass, N., & Scheel, D. (2000). Receptor-mediated increase in cytoplasmic free calcium required for activation of pathogen defense in parsley. *The Plant Cell, 12*(8), 1425-1440.

Castillon, A., Shen, H., & Huq, E. (2007). Phytochrome interacting factors: central players in phytochrome-mediated light signaling networks. *Trends in plant science, 12*(11), 514-521.

Cenik, C., Cenik, E. S., Byeon, G. W., Grubert, F., Candille, S. I., Spacek, D., ... & Ricci, E. (2015). Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome research, 25*(11), 1610-1621.

Chang, C., Wang, Y. F., Kanamori, Y., Shih, J. J., Kawai, Y., Lee, C. K., ... & Esashi, M. (2005). Etching submicrometer trenches by using the Bosch process and its application to the fabrication of antireflection structures. *Journal of micromechanics and microengineering, 15*(3), 580-585.

Chattopadhyay, S., Ang, L. H., Puente, P., Deng, X. W., & Wei, N. (1998). Arabidopsis bZIP protein HY5 directly interacts with light-responsive promoters in mediating light control of gene expression. *The Plant Cell, 10*(5), 673-683.

Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant Journal, 89*(4), 789-804.

Chisholm, S. T., Coaker, G., Day, B., & Staskawicz, B. J. (2006). Host-microbe interactions: shaping the evolution of the plant immune response. *Cell, 124*(4), 803-814.

Coolen, S., Proietti, S., Hickman, R., Davila Olivas, N. H., Huang, P. P., Van Verk, M. C., ... & Van Loon, J. J. (2016). Transcriptome dynamics of Arabidopsis during sequential biotic and abiotic stresses. *The Plant Journal, 86*(3), 249-267.

Cunnac, S., Chakravarthy, S., Kvitko, B. H., Russell, A. B., Martin, G. B., & Collmer, A. (2011). Genetic disassembly and combinatorial reassembly identify a minimal functional repertoire of type III effectors in *Pseudomonas syringae*. *Proceedings of the National Academy of Sciences*, *108*(7), 2975-2980.

Dangl, J. L., Dietrich, R. A., & Richberg, M. H. (1996). Death don't have no mercy: cell death programs in plant-microbe interactions. *The Plant Cell*, *8*(10), 1793.

Day, B., Dahlbeck, D., Huang, J., Chisholm, S. T., Li, D., & Staskawicz, B. J. (2005). Molecular basis for the RIN4 negative regulation of RPS2 disease resistance. The *Plant Cell*, *17*(4), 1292-1305.

de Torres Zabala, M., Bennett, M. H., Truman, W. H., & Grant, M. R. (2009). Antagonism between salicylic and abscisic acid reflects early host–pathogen conflict and moulds plant defence responses. *The Plant Journal*, *59*(3), 375-386.

DeYoung, B. J., & Innes, R. W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. *Nature immunology*, *7*(12), 1243-1249.

Eulgem, T., & Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. *Current opinion in plant biology*, *10*(4), 366-371.

Eulgem, T., Rushton, P. J., Robatzek, S., & Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. *Trends in plant science*, *5*(5), 199-206.

Feys, B. J., Benedetti, C. E., Penfold, C. N., & Turner, J. G. (1994). Arabidopsis mutants selected for resistance to the phytotoxin coronatine are male sterile, insensitive to methyl jasmonate, and resistant to a bacterial pathogen. *The Plant Cell, 6*(5), 751-759.

Fraley, C., & Raftery, A. E. (2006). Model-based microarray image analysis. *The Newsletter of the R Project, 6*(5), 60-63.

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., & Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences, 111*(6), 2367-2372.

Gallego-Bartolomé, J., Minguet, E. G., Marín, J. A., Prat, S., Blázquez, M. A., & Alabadí, D. (2010). Transcriptional diversification and functional conservation between DELLA proteins in Arabidopsis. *Molecular biology and evolution, 27*(6), 1247-1256.

Gimenez-Ibanez, S., Boter, M., Fernández-Barbero, G., Chini, A., Rathjen, J. P., & Solano, R. (2014). The bacterial effector HopX1 targets JAZ transcriptional repressors to activate jasmonate signaling and promote infection in Arabidopsis. *PLoS Biology, 12*(2), e1001792.

Glazebrook, J. (2001). Genes controlling expression of defense responses in Arabidopsis—2001 status. *Current opinion in plant biology, 4*(4), 301-308.

Goodman, R. N., & Novacky, A. J. (1994). *The hypersensitive reaction in plants to pathogens: a resistance phenomenon.* American Phytopathological Society (APS).

Griebel, T., & Zeier, J. (2008). Light regulation and daytime dependency of inducible plant defenses in Arabidopsis: phytochrome signaling controls systemic acquired resistance rather than local defense. *Plant physiology, 147*(2), 790-801.

Guan, Y., Chen, M., Ma, Y., Du, Z., Yuan, N., Li, Y., ... & Zhang, Y. (2020). Whole-genome and time-course dual RNA-Seq analyses reveal chronic pathogenicity-related gene dynamics in the ginseng rusty root rot pathogen *Ilyonectria robusta. Scientific reports, 10*(1), 1586-1603.

Hadwiger, L. A., & Tanaka, K. (2017). Non-host resistance: DNA damage is associated with SA signaling for induction of PR genes and contributes to the growth suppression of a pea pathogen on pea endocarp tissue. *Frontiers in Plant Science, 8*(446), 1-12.

Helft, L., Thompson, M., & Bent, A. F. (2016). Directed evolution of FLS2 towards novel flagellin peptide recognition. *PloS one, 11*(6), e0157155.

Higashi, K., Ishiga, Y., Inagaki, Y., Toyoda, K., Shiraishi, T., & Ichinose, Y. (2008). Modulation of defense signal transduction by flagellin-induced WRKY41 transcription factor in *Arabidopsis thaliana*. *Molecular Genetics and Genomics, 279*(3), 303-312.

Hoang, T. V., Vo, K. T. X., Hong, W. J., Jung, K. H., & Jeon, J. S. (2018). Defense response to pathogens through epigenetic regulation in rice. *Journal of Plant Biology, 61*(1), 1-10.

Hou, X., Lee, L. Y. C., Xia, K., Yan, Y., & Yu, H. (2010). DELLAs modulate jasmonate signaling via competitive binding to JAZs. *Developmental cell, 19*(6), 884-894.

Huot, B., Yao, J., Montgomery, B. L., & He, S. Y. (2014). Growth–defense tradeoffs in plants: a balancing act to optimize fitness. *Molecular plant, 7*(8), 1267-1287.

Ji, Z., Ji, C., Liu, B., Zou, L., Chen, G., & Yang, B. (2016). Interfering TAL effectors of *Xanthomonas oryzae* neutralize R-gene-mediated plant disease resistance. *Nature communications, 7*(1), 1-9.

Ji, Z., Wang, C., & Zhao, K. (2018). Rice routes of countering *Xanthomonas oryzae*. *International Journal of Molecular Sciences, 19*(10), 3008.

Jiang, S., Yao, J., Ma, K. W., Zhou, H., Song, J., He, S. Y., & Ma, W. (2013). Bacterial effector activates jasmonate signaling by directly targeting JAZ transcriptional repressors. *PLoS Pathogens, 9*(10), e1003715.

Jones, J. D., & Dangl, J. L. (2006). The plant immune system. *Nature, 444*(7117), 323-329.

Kang, H. G., & Dingwell, J. B. (2008). Effects of walking speed, strength and range of motion on gait stability in healthy older adults. *Journal of biomechanics, 41*(14), 2899-2905.

Kang, H. G., & Singh, K. B. (2001). Characterization of salicylic acid-responsive, Arabidopsis Dof domain proteins: overexpression of OBP3 leads to growth defects. *The Plant Journal, 21*(4), 329-339.

Katagiri, F., & Chua, N. H. (1992). Plant transcription factors: present knowledge and future challenges. *Trends in Genetics, 8*(1), 22-27.

Kenton, P., Mur, L. A., Atzorn, R., Wasternack, C., & Draper, J. (1999). (—)-Jasmonic acid accumulation in Tobacco hypersensitive response lesions. *Molecular plant-microbe interactions, 12*(1), 74-78.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., Van Der Lee, R., ... & Baranasic, D. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic acids research, 46*(D1), D260-D266.

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology, 37*(8), 907-915.

Kim, M. G., Geng, X., Lee, S. Y., & Mackey, D. (2009). The Pseudomonas syringae type III effector AvrRpm1 induces significant defenses by activating the Arabidopsis nucleotide-binding leucine-rich repeat protein RPS2. *The Plant Journal, 57*(4), 645-653.

Kindgren, P., Norén, L., Lopez, J. D. D. B., Shaikhali, J., & Strand, Å. (2012). Interplay between Heat Shock Protein 90 and HY5 controls PhANG expression in response to the GUN5 plastid signal. *Molecular Plant, 5*(4), 901-913.

Leivar, P., & Monte, E. (2014). PIFs: systems integrators in plant development. *The Plant Cell, 26*(1), 56-78.

Leivar, P., & Quail, P. H. (2011). PIFs: pivotal components in a cellular signaling hub. *Trends in plant science, 16*(1), 19-28.

Lewis, L. A., Polanski, K., de Torres-Zabala, M., Jayaraman, S., Bowden, L., Moore, J., ... & Kulasekaran, S. (2015). Transcriptional dynamics driving MAMP-triggered immunity and pathogen effector-mediated immunosuppression in Arabidopsis leaves following infection with Pseudomonas syringae pv tomato DC3000. *The Plant Cell, 27*(11), 3038-3064.

Li, B., Meng, X., Shan, L., & He, P. (2016). Transcriptional regulation of pattern-triggered immunity in plants. *Cell Host & Microbe, 19*(5), 641-650.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Li, K., Yu, R., Fan, L. M., Wei, N., Chen, H., & Deng, X. W. (2016). DELLA-mediated PIF degradation contributes to coordination of light and gibberellin signalling in Arabidopsis. *Nature communications, 7*(1), 1-11.

Li, M., Ma, X., Chiang, Y. H., Yadeta, K. A., Ding, P., Dong, L., ... & Shen, Q. H. (2014). Proline isomerization of the immune receptor-interacting protein RIN4 by a cyclophilin inhibits effector-triggered immunity in Arabidopsis. *Cell host & microbe, 16*(4), 473-483.

Ligges, U., & Mächler, M. (2002). Scatterplot3d – an R package for visualizing multivariate data. *Journal of Statistical Software 8*(11), 1–20

Liu, J., Elmore, J. M., Lin, Z. J. D., & Coaker, G. (2011). A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Cell host & microbe*, *9*(2), 137-146.

Liu, L., Sonbol, F. M., Huot, B., Gu, Y., Withers, J., Mwimba, M., ... & Dong, X. (2016). Salicylic acid receptors activate jasmonic acid signalling through a non-canonical pathway to promote effector-triggered immunity. *Nature communications*, *7*(1), 1-10.

Liu, X., Liu, R., Li, Y., Shen, X., Zhong, S., & Shi, H. (2017). EIN3 and PIF3 form an interdependent module that represses chloroplast development in buried seedlings. *The Plant Cell, 29*(12), 3051-3067.

Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R., & Dangl, J. L. (2003). Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell, 112*(3), 379-389.

Malik, S., & Van der Hoorn, R. A. (2016). Inspirational decoys: a new hunt for effector targets. *The New Phytologist, 210*(2).

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal, 17*(1), 10-12.

McLeay, R. C., & Bailey, T. L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC bioinformatics*, *11*(1), 165.

Mine, A., Seyfferth, C., Kracher, B., Berens, M. L., Becker, D., & Tsuda, K. (2018). The defense phytohormone signaling network enables rapid, high-amplitude transcriptional reprogramming during effector-triggered immunity. *The Plant Cell, 30*(6), 1199-1219.

Mine, A., Seyfferth, C., Kracher, B., Berens, M. L., Becker, D., & Tsuda, K. (2018). The defense phytohormone signaling network enables rapid, high-amplitude transcriptional reprogramming during effector-triggered immunity. *The Plant Cell, 30(*6), 1199-1219.

Mishina, T. E., & Zeier, J. (2007). Pathogen-associated molecular pattern recognition rather than development of tissue necrosis contributes to bacterial induction of systemic acquired resistance in Arabidopsis. *The Plant Journal, 50*(3), 500-513.

Murrell, P. (2018). *R graphics.* Boca Raton, FL: CRC Press.

Nagoshi, E., Saini, C., Bauer, C., Laroche, T., Naef, F., & Schibler, U. (2004). Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell, 119*(5), 693-705.

Nath, V. S., Mishra, A. K., Kumar, A., Matoušek, J., & Jakše, J. (2019). Revisiting the role of transcription factors in coordinating the defense response against citrus bark cracking viroid infection in commercial hop (*Humulus Lupulus* L.). *Viruses, 11*(419), 1-19.

Nazar, R., Iqbal, N., & Khan, N. A. (Eds.). (2017). *Salicylic Acid: A Multifaceted Hormone*. Singapore: Springer.

Nose, T., Waseda, T., Kodaira, T., & Inoue, J. (2020). Satellite-retrieved sea ice concentration uncertainty and its effect on modelling wave evolution in marginal ice zones. *The Cryosphere, 14*(6), 2029-2052.

Nueda, M. J., Tarazona, S., & Conesa, A. (2014). Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics, 30*(18), 2598-2602.

Paik, H., Chen, Z., Lochocki, E., Seidner H, A., Verma, A., Tanen, N., ... & Brützam, M. (2017). Adsorption-controlled growth of La-doped BaSnO3 by molecular-beam epitaxy. *APL Materials, 5*(11), 116107.

Paik, I., Kathare, P. K., Kim, J. I., & Huq, E. (2017). Expanding roles of PIFs in signal integration from multiple processes. *Molecular plant, 10*(8), 1035-1046.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols, 11*(9), 1650.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology, 33*(3), 290-295.

Ranf, S., Eschen-Lippold, L., Pecher, P., Lee, J., & Scheel, D. (2011). Interplay between calcium signalling and early signalling elements during defence responses to microbe-or damage-associated molecular patterns. *The Plant Journal, 68*(1), 100-113.

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology, 11*(3), 1-9.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140.

Sarris, P. F., Duxbury, Z., Huh, S. U., Ma, Y., Segonzac, C., Sklenar, J., ... & Wirthmueller, L. (2015). A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell, 161*(5), 1089-1100.

Schechter, L. M., Vencato, M., Jordan, K. L., Schneider, S. E., Schneider, D. J., & Collmer, A. (2006). Multiple approaches to a complete inventory of *Pseudomonas syringae* pv. *tomato* DC3000 type III secretion system effector proteins. *Molecular plant-microbe interactions, 19*(11), 1180-1192.

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal, 8(*1), 289.

Smakowska, E., Kong, J., Busch, W., & Belkhadir, Y. (2016). Organ-specific regulation of growth-defense tradeoffs by plants. *Current opinion in plant biology, 29*, 129-137.

Sohn, K. H., Saucet, S. B., Clarke, C. R., Vinatzer, B. A., O'Brien, H. E., Guttman, D. S., & Jones, J. D. (2012). HopAS1 recognition significantly contributes to Arabidopsis nonhost resistance to *Pseudomonas syringae* pathogens. *New Phytologist, 193*(1), 58-66.

Spies, D., Renz, P. F., Beyer, T. A., & Ciaudo, C. (2019). Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Briefings in bioinformatics, 20*(1), 288-298.

Spoel, S. H., Johnson, J. S., & Dong, X. (2007). Regulation of tradeoffs between plant defenses against pathogens with different lifestyles. *Proceedings of the National Academy of Sciences, 104*(47), 18842-18847.

Stephenson, P. G., Fankhauser, C., & Terry, M. J. (2009). PIF3 is a repressor of chloroplast development. *Proceedings of the National Academy of Sciences, 106*(18), 7654-7659.

Stuart, L. M., Paquette, N., & Boyer, L. (2013). Effector-triggered versus pattern-triggered immunity: how animals sense pathogens. *Nature Reviews Immunology, 13*(3), 199-206.

Sun, L., & Zhang, J. (2020). Regulatory role of receptor-like cytoplasmic kinases in early immune signaling events in plants. *FEMS Microbiology Reviews, fuaa035,* 1–12

Tao, Y., Xie, Z., Chen, W., Glazebrook, J., Chang, H. S., Han, B., ... & Katagiri, F. (2003). Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *The Plant Cell*, *15*(2), 317-330.

Team, R. C., & Worldwide, C. (2002). The R stats package. R Foundation for Statistical Computing, Vienna, Austria: Available from: http://www. R-project. org.

Thrall, P. H., & Burdon, J. J. (2003). Evolution of virulence in a plant host-pathogen metapopulation. *Science, 299*(5613), 1735-1737.

Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J., & Katagiri, F. (2009). Network properties of robust immunity in plants. *PLoS Genet, 5*(12), e1000772.

Ülker, B., & Somssich, I. E. (2004). WRKY transcription factors: from DNA binding towards biological function. *Current opinion in plant biology, 7*(5), 491-498.

Van Der Biezen, E. A., & Jones, J. D. (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trends in biochemical sciences, 23*(12), 454-456.

Verbist, B. M., Thys, K., Reumers, J., Wetzels, Y., Van der Borght, K., Talloen, W., ... & Thas, O. (2015). VirVarSeq: a low-frequency virus variant detection pipeline for Illumina sequencing using adaptive base-calling accuracy filtering. *Bioinformatics*, *31*(1), 94-101.

Wehrens, R., & Buydens, L. M. (2007). Self-and super-organizing maps in R: the Kohonen package. *Journal of Statistical Software*, *21*(5), 1-19.

Wehrens, R., & Kruisselbrink, J. (2018). Flexible Self-Organising Maps in kohonen 3.0. *Journal of Statistical Software, 87*(7), 1-18.

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180-185.

Yanagisawa, S. (2002). The Dof family of plant transcription factors. *Trends in plant science, 7*(12), 555-560.

Yang, L., Teixeira, P. J. P. L., Biswas, S., Finkel, O. M., He, Y., Salas-Gonzalez, I., ... & Dangl, J. L. (2017). *Pseudomonas syringae* type III effector HopBB1 promotes host transcriptional repressor degradation to regulate phytohormone responses and virulence. *Cell host & microbe, 21*(2), 156-168.

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR), 52*(1), 1-38.

Zhang, X., Huai, J., Shang, F., Xu, G., Tang, W., Jing, Y., & Lin, R. (2017). A PIF1/PIF3-HY5-BBX23 transcription factor cascade affects photomorphogenesis. *Plant Physiology, 174*(4), 2487-2500.

Zhang, Y., Tian, L., Yan, D. H., & He, W. (2018). Genome-wide transcriptome analysis reveals the comprehensive response of two susceptible poplar sections to *Marssonina brunnea* infection. *Genes, 9*(3), 154.

Zhou, Z., Wu, Y., Yang, Y., Du, M., Zhang, X., Guo, Y., ... & Zhou, J. M. (2015). An Arabidopsis plasma membrane proton ATPase modulates JA signaling and is exploited by the *Pseudomonas syringae* effector protein AvrB for stomatal invasion. *The Plant Cell, 27*(7), 2032-2041.

Zhu, Q. H., Stephen, S., Kazan, K., Jin, G., Fan, L., Taylor, J., ... & Wang, M. B. (2013). Characterization of the defense transcriptome responsive to *Fusarium oxysporum*-infection in Arabidopsis using RNA-seq. *Gene, 512*(2), 259-266.

Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E. J., Jones, J. D., Felix, G., & Boller, T. (2004). Bacterial disease resistance in Arabidopsis through flagellin perception. *Nature*, *428*(6984), 764-767.

Zou, J., Rodriguez-Zas, S., Aldea, M., Li, M., Zhu, J., Gonzalez, D. O., ... & Clough, S. J. (2005). Expression profiling soybean response to *Pseudomonas syringae* reveals new defense-related genes and rapid HR-specific downregulation of photosynthesis. *Molecular plant-microbe interactions*, *18*(11), 1161-1174.